A systematic framework for resource use measurement in the economic evaluation of individual-level physical activity and sedentary behaviour interventions

Madeleine Cochrane

A thesis submitted in partial fulfilment of the requirements of

Liverpool John Moores University for the degree of Doctor of Philosophy

July 2020

Table of Contents

Abstract

There is an increasing demand for economic evidence on physical activity (PA) and sedentary behaviour (SB) interventions which can prevent noncommunicable disease (NCD). Trials assessing the impact of PA and SB interventions rarely collect and present economic data alongside their effectiveness results. The overarching aim of this PhD was to develop a multidisciplinary and pragmatic framework to support researchers carrying out trial-based economic evaluations of individual-level PA and SB interventions. The nature of this PhD project presented me with the opportunity to train in a new discipline, health economics, and draw on my existing discipline-specific knowledge in anthropology and public health in order to make an interdisciplinary contribution to the field of public health economics applied to trials. In particular, my knowledge in anthropology, the study of human society and its complexity, supported my documentation of the complexity involved in developing and implementing a standardised pragmatic framework to the multidisciplinary field of economic evaluation in PA and SB. Complexity is reported throughout this thesis in the form of reflection boxes. The purpose of the reflections were to: (1) describe in detail the actions I took to develop a framework; and (2) explain why I believed these actions were fit for purpose. For the latter, I reflect on my prior knowledge in public health and anthropology, as well as on the complex detailed observations I made and informal multidisciplinary conversations I had during this PhD.

In order to design an initial framework and to test the practicability of it, I carried out three key studies. Study one was a systematic review which aimed to identify how authors of existing economic evaluations of individual-level PA and SB interventions have addressed key methodological challenges associated with the conduct of economic evaluations in public health. In summary, the review showed that there are marked methodological differences in existing studies. Nevertheless, good quality methods were identified and drawn on in order to develop the initial framework. The initial framework comprised of 16 items along with guidance on how these items could be applied in the context of PA and SB. Study two involved applying the initial framework to a ‘real world’ PA trial and reflecting on its practicability. Study three was carried out concurrently with study two and involved applying the initial framework to a ‘real world’ SB trial, also to see how the framework performed in practice. Narrative synthesis methods were used to bring together the key learnings and reflections from studies 1-3. The narrative synthesis shed light on how my interdisciplinary knowledge and experience could improve the procedures for identifying and measuring resource use within PA and SB trials. My revised systematic framework incorporates existing tools from the multiple research fields in which PA and SB cut across, namely public health, exercise science, behavioural science, anthropology and trial methodology. The final framework, presented in the form of a standard operating procedure (SOP), is recommended for use in trial units to support early career health economists to make and communicate decisions around the measurement of resource use in complex individual-level PA and SB trials.

Declaration

No proportion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. This research programme had a methodological aim. However, two of the empirical studies from which some of the demographic and economic data in this thesis derived from were conducted as part of two collaborative trials called Co-PARS and SLaMM. Data from the Co-PARS trial was collected by two students: Madeleine Cochrane and another full-time PhD student (Ben Buckley). Madeleine Cochrane's contribution was that she planned, input, analysed and presented all economic data for the three comparator groups for the Co-PARS trial without the assistance of Ben Buckley. In addition, at the 6-month follow up time point Madeleine Cochrane collected all economic and non-economic data for the Co-PARS trial for two of the comparator groups (Co-PARS intervention group and usual care group) without the assistance of Ben Buckley. For the SLaMM trial, Madeleine Cochrane contributed to the design and delivery of the two interventions and trial alongside one full-time PhD student (Abby Morris) and one part-time MPhil student (David Gavin). Madeleine Cochrane also collected all economic and non-economic data alongside Abby Morris and David Gavin, at the baseline and 12-week follow up time points for both comparator groups. In addition, Madeleine Cochrane planned, input, analysed and presented all economic data for the two comparator groups without the assistance of Abby Morris and David Gavin.

Peer-reviewed publication

Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. (2019). Systematic review of the methods used in economic evaluations of targeted PA and SB interventions. *Social Science & Medicine, 232, 156-167*. doi: https://doi.org/10.1016/j.socscimed.2019.04.040

Conference communications

Conference	**Presentation title, authors, date and location**
International Society for PA and Health *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. October 2018, London
International PhD Symposium *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions.

	Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. July 2018, Liverpool
Faculty of Science Research Day, LJMU *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. June 2018, Liverpool
Faculty of Science Research Day, LJMU *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. June 2018, Liverpool
Health Economics Training Group Seminar Series, Deakin University *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. March 2018, Melbourne, Australia
UK Society for Behavioural Medicine *Oral*	Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F.

	December 2017, Liverpool
Research Institute for Sport and Exercise Sciences (RISES) 20 Year Celebration Seminar Series *Oral*	An evidence-based workplace intervention for call agents: a 'real-world' project. *Economic cost of SB* Cochrane, M., Morris, A.S., Gavin, D., Graves, L. E. F. November 2017, Liverpool
Health Economic Seminar Series, Liverpool School of Tropical Medicine *Oral*	Protocol for a systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. December 2017, Liverpool
Graduate School Annual PhD Conference, LJMU *Poster*	Protocol for a systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. May 2017, Liverpool
Faculty of Science Research Day, LJMU *Poster*	Protocol for a systematic review of the methods used in economic evaluations of individual-level PA and SB interventions. Cochrane, M., Watson, P. M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin,A., Graves, L. E. F. May 2017, Liverpool

Acknowledgements

Academic: A huge thank you to my brilliant supervisory team: Dr Lee Graves, Dr Paula Watson, Dr Hannah Timpson and Dr Brendan Collins. All of you have invested a lot of time in my PhD project, shared your great ideas and expertise with me, and have encouraged me consistently over the last three years- I really appreciate this! Thank you also to Dr Alan Haycox, Dr Antony Martin and Lisa Jones for offering your time and expertise particularly at the early stages of the PhD, which contributed to me getting my first author publication (which I was delighted about!). A big thank you to the academic staff and PhD students at the PA Exchange who have created a supportive PhD environment over the last three years. A special thanks to Dr Abby Morris, Dr Ben Buckley and David Gavin for being a great team to work alongside- working hard but still remembering to have fun.

Personal: Thank you to my parents, siblings, aunties and wider family, as well as my Durham and Caravan friends. You have all cheered me on and motivated me with your own success and resilience. I feel tremendously lucky to have such a positive, kind and hard-working group of people around me. In particular, thank you to Geoff for your encouragement and thoughtfulness. Importantly, a special thanks to my mum (Frances) and sister (Georgina) who have been my lifelong cheerleaders and role models from the very start. Your enthusiasm, optimism and reassurance has helped me get where I am today- Merci beaucoup!

List of Tables

List of Figures

Abbreviations

Physical Activity	PA
Sedentary Behaviour	SB
Noncommunicable disease	NCD
Cost-utility analysis	CUA
Cost-effectiveness analysis	CEA
Cost-consequence analysis	CCA
Cost-benefit analysis	CBA
National Institute for Health and Care Excellence	NICE
Quality-adjusted life year	QALY
Health-related quality of life	HRQoL
European quality of life- five dimensions	EQ-5D
Incremental cost-effectiveness ratio	ICER
Cost-effectiveness acceptability curve	CEAC
Willingness to pay	WTP
95% Confidence Interval	95% CI
Exercise Referral Scheme	ERS
Exercise Referral Practitioner	ERP
Co-developed PA on Referral Scheme	Co-PARS
Sit Less and Move More intervention	SLaMM
Randomised controlled trial	RCT
Database of Instruments for Resource Use Measurement	DIRUM
Consolidated Health Economic Evaluation Reporting Standards	CHEERS
Preferred Reporting Items for Systematic Reviews and Meta Analyses	PRISMA
Consolidated standards of reporting trials	CONSORT
Template for intervention description and replication	TIDieR
Pragmatic Explanatory Continuum Indicator Summary- version 2	PRECIS-2
Personal Social Services Research Unit	PSSRU
Client Service Receipt Inventory	CSRI
Health Resource Groups	HRG
Office of National Statistics	ONS
Intention-to-treat	ITT
Analysis of covariance	ANCOVA
Minimum Clinically Important Difference	MCID
Standard Deviation	SD
Standard Error	SE
Coronary Heart Disease	CHD
Cardiovascular disease	CVD
Type 2 diabetes	T2D
World Health Organisation	WHO

Chapter 1: Introduction

1.1. Definitions for PA and SB

Physical activity (PA) has been formally defined as “any bodily movement produced by skeletal muscles that results in energy expenditure” (Caspersen et al., 1985). This definition implies that PA is not necessarily planned, structured or repetitive. It indicates that PA can be performed as part of people’s daily life activities such as through workplace and household activities, as well as exercise and sport. In the PA literature, energy used up through PA is classified in units known as metabolic equivalent time (MET). METs have been defined as “the amount of oxygen consumed during PA” (Scholes, 2017). The Compendium for Physical Activities categorises PA into three levels based on the amount of METs used. These categories include light-intensity PA (1.6-2.9 METs), moderate-intensity PA (2-5.9 METs) and vigorous-intensity PA (≥6 METs) (Ainsworth et al., 2011). More recently, the terminology for bodily movement has been extended to incorporate two low energy behaviours: sedentary behaviour (SB) and sleep. SB has been defined as “any waking behaviour characterized by an energy expenditure ≤1.5 METs while in a sitting or reclining posture” (Sedentary Behaviour Research Network, 2012). Sleep has been classified as using less than 1 MET (Tremblay et al., 2017).

1.2. Why intervention is needed for PA and SB

1.2.1. Evidence on the benefits of PA and SB

Since the industrial revolution, there has been a rapid increase in the development of new technology. In turn this increase in technology has led to a reduction in human energy expenditure since new technology has reduced the heavy labour and effort required to perform day-to-day activities (Hallal et al., 2012). One of the consequences of modern civilisation is an increased risk of noncommunicable disease (NCD). This is because key biological mechanisms in the human body are maladapted to a lifestyle that is less active (Booth et al., 2008). As a result, it has been claimed that there is overwhelming evidence to assert that physical inactivity is one the biggest public health challenges of the 21st Century (Blair, 2009). NCDs are important as they are the main cause for death and disability worldwide with incidence rates forecast to increase considerably (WHO, 2014). NCDs are often recognisable by their long-term nature and include conditions such as cardiovascular disease (CVD), type II diabetes (T2D), cancer, dementia and depression. Physical inactivity is one of the four main lifestyle risk factors, which has and continues to contribute to the global rise in NCDs and premature mortality (World Health Organisation, 2013). As a result, physical inactivity is the fourth leading cause of death globally (Kohl et al., 2012) and is attributable to one in six global deaths (Lee et al., 2012). Physical inactivity has been associated with a number of major NCDs including dementia, coronary heart disease (CHD), T2D, stroke, breast and colon cancer, and depression (Andersen et al., 2016, Department of Health, 2004, Trueman and Anokye, 2013).

The most recent guidelines on PA from the UK emphases that: (1) any amount of PA can lead to health benefits; and (2) all age groups can benefit from PA (Gibson-Moore, 2019). Levels of PA below the current recommendations have even demonstrated important health benefits. A key review looking at the health effects of light-intensity PA for adults and older adults measured through objective measures, found that light-intensity PA was beneficially associated with obesity,

cardiometabolic markers and mortality (Füzéki et al., 2017). Furthermore, low doses of MVPA incorporated into an older persons daily life can reduce their risk of mortality by 22% (Hupin et al., 2015). There is also evidence indicating that a single bout of resistance PA can have important health benefits for adults including a lowering effect on blood pressure which can last up to 24 hours (Casonatto et al., 2016). In terms of mental health, a cross-sectional study from the USA with data from 1.2 million individuals found that individuals who reported exercising in the past month compared to those who did not, spent 1.49 days less per month in poor mental health. Furthermore, the authors found that all types of exercise were associated with a reduction in poor mental health (Chekroud et al., 2018). In terms of children and young people, epidemiological evidence indicate that children who develop a physically active lifestyle very early on in childhood are more likely to be active across the life course (Telama et al., 2014). There are also immediate benefits for children and young people, for instance, a large systematic review examining the evidence on objectively measure PA in 5-17 year olds from 162 studies, found that all forms of PA in terms of sporadic, bouts and continuous PA, are beneficial for children and young people's health (Poitras et al., 2016). More specifically, the review reported an association between PA and physical, psychological, social and cognitive benefits. Overall, the benefits of PA are widespread for all age groups and they do not just include physical and mental health benefits. For instance, PA is associated with improvements in learning and attainment for children as well as improvements in workplace productivity for adults (Gibson-Moore, 2019).

SB has also been identified as modifiable risk factor associated with NCD and premature mortality. Importantly, SB has been identified as having a negative impact on health, independent of moderate-to-vigorous intensity PA (Katzmarzyk et al., 2009, Hamilton et al., 2008, Buckley et al., 2015). A key meta-analysis found that adults who were sedentary for prolonged periods were still at increased risk of mortality and morbidity, regardless of whether they were meeting the weekly guidelines for PA (Owen et al., 2010, Ekelund et al., 2016). The study found that in order to attenuate the risks associated with SB, adults would need to participate in 60 minutes of 'moderate-intensity' PA per day. As over a third (35%) of females and a quarter (26%) of males in high-income countries do not presently meet the recommended weekly guidelines of 150 minutes of PA per week (WHO, 2018b) then a daily target of 60 minutes, which equates to a weekly target of 420 minutes, is unlikely to be attained. A more recent meta-analysis involving over one million participants showed that high levels of total sitting and especially TV viewing time, are associated with an increased risk of all-cause mortality, CVD mortality and incidence of T2D (Patterson et al., 2018). Sitting for prolonged periods has also been associated with a decline in metabolic health. A large study from Australia found that for adults without a diagnosis of diabetes, who self-reported high-levels of TV viewing time, were also more likely to have undiagnosed abnormal glucose metabolism (Dunstan et al., 2004). Similarly, a large systematic review drawing on 235 studies from 71 countries, found an association between TV viewing for children and young people, and unfavourable cardiometabolic risk scores, body composition and behavioural conduct (Carson et al., 2016).

1.2.2. Recommended guidelines for PA and SB

The terms 'moderate' and 'vigorous' are drawn on in the international guidelines for PA which aim to recommend how much PA adults should achieve per week in order to reduce their risk of NCD and pre-mature mortality (WHO, 2018b). Currently, the guidelines state that adults aged 18-64 should aim to achieve at least 150 minutes of 'moderate-intensity' PA per week or 75 minutes of 'vigorous-intensity' PA per week, or the equivalent weekly volume through a combination of moderate-to-vigorous PA. These guidelines were based on key systematic reviews from the US and Canada (Department of Health and Human Services, 2008, Warburton et al., 2010). Though the reviews were published over a decade ago, the evidence derived from them remains relevant today. For instance, the reviews found that 150 minutes of 'moderate-intensity' PA per week is associated with substantial health benefits across a diverse range of adult populations. Furthermore, they found that it is the overall volume of 150 minutes of 'moderate-intensity' PA per week, which is important, as opposed to the type of activity or frequency of PA sessions. Similar health benefits are accumulated when overall PA volume equates to 150 minutes per week, meaning it does not matter if this amount is achieved through short 10-minute bouts of PA or through long continuous PA sessions. That said, the guidelines do go on to recommend that in order to experience more acute benefits, it is better to spread the 150 minutes across the week. This is because acute effects of PA such as improved mood, insulin sensitivity and fat metabolism, only last up to 24-48 hours after a bout of PA (Department of Health and Social Care, 2011). Guidance has been provided by experts in order to improve our understanding further for the terms 'moderate' and 'vigorous' activities. This is because METs is not a multidisciplinary measure, therefore the guidance states "moderate-intensity activities are those in which heart rate and breathing are raised, but it is possible to speak comfortably. Vigorous-intensity activities are those in which heart rate is higher, breathing is heavier, and conversation is harder" (O'Donovan et al., 2010). On the contrary, to PA guidelines, due to the underdeveloped evidence base for SB, there are no international guidelines for SB and many countries have not quantified guidelines (Stamatakis et al., 2019). For example, the UK's SB guidelines recommends that adults minimise the amount of time they are sedentary for extended periods but the guidelines do not quantify this (Department of Health and Social Care, 2011).

1.2.3. Current global and national prevalence for PA and SB

Globally, around one in four adults are not meeting the recommended guidelines for PA (WHO, 2018b). Nationally, the latest Health Survey for England reported similar findings while also noting that a greater proportion of females (42%) are not meeting the guidelines compared to males (34%) (Scholes, 2017). In terms of SB, early evidence from the US indicated that Americans reportedly spend 55% of their leisure time in sedentary pursuits, including watching TV (Matthews et al., 2008). While in Europe, early evidence on SB suggested that European's devote 40% of their leisure time to watching TV (Office for European Communities, 2003). In the context of the UK, the Health Survey for England (Scholes, 2017) also assessed SB levels and found that the same proportion (29%) of males and females are sedentary for six hours or more per day during their leisure time in the week. That said, at the weekend males were more likely to spend 6 or more hours of their leisure time per day being sedentary (40% vs 35% per day, respectively). The

difference between the findings from America and the UK may relate to the way SB was measured in the studies. The study by Matthew and colleagues uses objective measures, while the Health Survey for England draws on self-report measures. When objective and self-report measures have been compared in previous work, in the context of PA, the objective data has indicated that people may overestimate how active they are when they self-report their activity levels (Marteau, 2018, Scarborough et al., 2011). Overall, the current evidence base on PA and SB prevalence suggests that an important share of people's waking hours is spent using very little energy.

1.2.4. Costs associated with PA and SB

The high prevalence and long duration of NCDs means their impact is widespread. Not only do they impact on the individual but also on numerous groups and sectors in society including the individual's family, workplace, community and health sector. As a consequence, the global burden of NCD is substantial (WHO, 2018a). In the UK alone, NCDs are estimated to cost the economy £8.3 billion per annum (Gray et al. 2015; Department of Health, 2009). Inaction to invest in preventative interventions tackling detrimental levels of physical inactivity is expected to lead to greater levels of NCD, inequity, productivity losses and a continued overwhelming demand for costly curative healthcare services (Organisation for Economic Cooperation and Development, 2015). Conservative estimates from 2013, report that the global economic burden of physical inactivity was around $58.3 billion per annum, of which more than half of this burden fell on the public sector (Ding et al., 2016). In the UK, Public Health England reported that physical inactivity costs the UK economy around £7.4 billion per annum (Public Health England, 2014). In other high income countries, physical inactivity is estimated to account for 1.5-3% of the total direct healthcare expenditure (Oldridge, 2008). In England, in 2009/10 the direct cost of physical inactivity to the National Health Service (NHS) was estimated to be around £900 million (Townsend et al., 2015).

1.2.5. Determinants of PA and SB

As well as reducing NCD and pre-mature mortality, an increase in PA and reduction in SB has the potential to contribute to making people more productive in the workplace and making the world more sustainable. For example, poor health has been associated with higher rates of workplace absenteeism and presenteeism (Scarborough et al., 2011, Buckley et al., 2015). Furthermore, it is expected that the promotion of PA through active transport can help reduce fossil fuel use, which in turn will lead to clearer air, less congested roads and ultimately a healthier environment (World Health Organisation, 2018). As well as PA and SB affecting productivity and the environment, work conditions and the environment can also affects people's health. Key milestone publications have helped improved our understanding of this in terms of how fundamental economic, environmental and social conditions can determine health-behaviours such as PA and SB. In particular, there have been three key publications, which have raised awareness on the determinants of health, these include: the Ottawa Charter for health promotion (World Health Organisation, 1986); Dahlgren and Whitehead's (1991) socioecological model; and the Marmot Review (Marmot et al., 2010). Together, the reports illustrate that interventions need to be designed with an awareness of the prerequisites for health. In particular, the Marmot review draws on robust epidemiological evidence, which shows how that it is not just genetics, which determine our health and health

behaviour, but the type of environmental and societal structures in which we live also play a major role in determining our lifestyle behaviours and consequently our health status.

1.3. Evidence on effectiveness of PA and SB interventions

1.3.1. Complex interventions for PA and SB

The determinants of PA and SB can change over time and are interrelated; this makes low PA and high SB levels complex public health challenges. This complexity has been explained through a recent framework produced as part of the Global Action Plan on PA (Rutter et al., 2018). The framework provides a visual depiction of the complex nature of PA, which can be used to help design effective PA and SB interventions. The framework builds on Dahlgren and Whitehead's socioecological approach as it helps us understand how contextual factors such as political, social, cultural and economic factors, influence people's response to PA interventions. More specifically, the framework disaggregates the influencing factors in order to demonstrate how there are multiple sectors and stakeholders involved in tackling the global problem of inactivity. The authors argue that the framework demonstrates that it is inappropriate to try and increase PA through a single response; rather they state that an effective response involves multiple sectors and multiple components. The Medical Research Council in the UK has provided guidance in order to help researchers deal with the complex nature of behaviour change interventions when developing and evaluating their interventions (Craig et al., 2008). The guidance is highly cited and due to be updated in order to incorporate recent methodological developments which includes complex systems thinking and natural experiments (Skivington et al., 2018). As the most recent version of the guidance is not yet published, key messages and concepts from 2008's guidance explain why many PA and SB interventions conducted over the last decade are multi-component, target different levels of the socioecological model, are evaluated through experimental designs, and assess multiple outcomes.

1.3.2. Effectiveness data for PA and SB interventions

International experts in PA have identified seven types of strategies for increasing PA, which they claim have worldwide applicability and are supported by good evidence (International Society for Physical Activity and Health, 2012). In summary, the seven strategies include:

1) Implementing whole school approaches to PA;
2) Creating transport systems which enable 'active transport';
3) Designing a built environment which provides opportunities for recreational PA and thus reduces people's chance of sitting for prolonged periods;
4) Encouraging primary and secondary healthcare professionals to prescribe PA as a form of 'medicine' for NCD prevention;
5) Raising awareness of PA benefits to the public through mass media campaigns;
6) Using key settings such as local governments, schools and workplaces to integrate PA promotion approaches across the whole community;
7) Encouraging participation in sport across the life span.

These seven strategies support the argument that an active lifestyle should be promoted at all stages of the life course. For example, although PA promotion has the greatest potential in the

early years, it also has an important role in supporting healthy workplaces and ageing (World Health Organisation, 2013). Traditionally, experts in PA have classified PA approaches into three broader strategies: community-wide mass media campaigns and informational approaches; individual-level behavioural and social support approaches; and community-wide environmental and policy approaches (Heath et al., 2012). Much of the existing evidence on the effectiveness of PA and SB interventions comes from the evaluation of individual-level behavioural and social support interventions; this is due to the challenge of measuring PA outcomes at the population-level. That said, the strength of these types of multicomponent individual-level interventions is that they can be developed and piloted in various settings before being scaled up to a community-wide and policy level (Craig et al., 2008).

In particular, there has much interest in individual-level behavioural and social support interventions, which aim to help individuals to incorporate PA into their daily routine. One reason for this is because early evidence has suggested that interventions which focus on lifestyle PA through the production of tailored activity plans are better value for money than supervised structured exercise programmes (Sevick et al., 2000). More recently, the evidence from a key meta-analysis, which reviewed the evidence on behaviour change techniques for PA, supported the evidence for tailored PA lifestyle interventions. More specifically, the meta-analysis found that goal setting, self-monitoring and person-centre methods (e.g. motivational interviewing and social determination theory) are effective techniques (Samdal et al., 2017). Compared to the PA literature, evidence on the effectiveness of SB interventions is less advanced. Overall, much of the evidence is limited to small non-powered and low quality studies (Shrestha et al., 2016). As office workers are one of the most sedentary populations (Clemes et al., 2014) SB interventions have typically been set in the workplace. A recent systematic review of the evidence on workplace SB interventions found that the existing evidence indicates that multicomponent behaviour change and environmental approaches are most effective for reducing workplace SB (Chu et al., 2016). That said, there remains a lack of evidence on the effectiveness of SB interventions that derives from RCTs (Gardner et al., 2016). Furthermore, there is a recognised need for more UK-based evidence on the effectiveness on workplace SB interventions (O'Connell et al., 2015) as well as evidence from a variety of workplaces (Mackenzie et al., 2015).

1.4. Economic evaluations in public health

1.4.1. Existing evidence on value for money

Epidemiological evidence on the benefits of PA from a range of low-income, middle-income and high-income countries, have led to PA being identified as a low cost global strategy for reducing mortality and CVD amongst adults aged 30-70 years (Lear et al., 2017). That said, evidence on which PA and SB interventions offer the best value for money is lacking. Not enough is known about the economic viability of PA and SB interventions and their potential for reducing future costs to health and social care. It is crucial that evaluations in this field consider economic and public health outcomes and costs to the individual (Anokye et al., 2014). One reason for the lack in evidence is because economic evaluations are rarely embedded in trials which evaluate the effectiveness of PA and SB interventions. Reviews from the UK, which have assessed the value for money of public health interventions, inclusive of a range of PA interventions, concluded that the

existing evidence on public health interventions indicates that they represent good value for money (Owen et al., 2012, Owen et al., 2017, Wanless, 2004). Wanless (2004) argues that the assessment of the cost-effectiveness of public health interventions alongside effectiveness evaluations should be routine practice. Nonetheless, a more recent review has concluded that not all preventative interventions are cost-effective and furthermore some cost-effectiveness results are sensitive to the methodological approaches used, as well as the choice in comparator groups, costs and assumptions (Owen and Fischer, 2019).

1.4.2. Fundamental economic concepts

The discipline of economics is underpinned by the concept of scarcity. More specifically, economists are interested in a concept known as opportunity cost. Opportunity cost relates to the idea that resources (e.g. materials and people's time) are limited and in most circumstances can only be used for one course of action at one point in time. Opportunity cost represents the consequence of allocating resources to one particular course of action instead of the next best alternative. That is to say, opportunity cost is the value of the benefits that could have been gained by choosing the next best alternative course of action instead (Morris et al., 2012). Edwards and McIntosh (2019) summarise three additional concepts which underpin the study of health economics, these include: allocative efficiency, technical efficiency and equity. Edwards and McIntosh (2019) explain these concepts in the context of public health decision-making within a public sector/ government economy:

- Allocative efficiency aims to consider how society can maximise society's welfare and asks: what public health goods, services and environments should be produced by society?
- Technical efficiency aims to consider how levels of input relate to levels of output and asks: how should public health goods, services and environments be produced?
- Equity aims to consider who to produce something for and asks: how should public health goods, services and environments be distributed across society?

1.4.3. Overview of economic evaluation approaches

Health economics is the study of how society allocates scarce healthcare resources. Economic evaluation is a key part of health economics as it compares the costs and effects of alternative courses of action (Drummond et al., 2015a). The principles and practice of health economics and economic evaluation can be applied to public health through the study of how society uses scarce resources to prevent ill health, promote healthy lifestyles and reduce inequalities (Edwards and McIntosh, 2019). There are four established approaches to economic evaluation, these include: cost effectiveness analysis (CEA), cost-utility analysis (CUA), cost-benefit analysis (CBA) and cost-consequence analysis (CCA). The main difference between the techniques is the way in which they incorporate and value outcomes. For instance, the main difference between conducting CEA or CBA in practice, relates to: (1) the number of outcomes included in the analysis; and (2) whether outcomes are monetised or reported in natural units.

In terms of CEA, this technique only compares aggregate costs to a single outcome measure. Furthermore, the chosen outcome measure is usually reported in natural units and not assigned a

monetary value. CUA is a variant of CEA as it also compares a single outcome measure to aggregate costs. However, CUA is unique as it uses preference-based outcome measures, which have preference weights attached to the various possible states of the outcome which means it is possible to rank and meaningfully compare outcomes. By contrast, CBA incorporate several outcome measures and CBA is not restricted to reporting outcome measures in their natural units or as a preference weight. CBA involves assigning monetary values to the different outcome measures. CBA presents a summary ratio statistic called the cost-benefit ratio which is the aggregate monetary value of all costs compared to the aggregate monetary value of all outcomes. Lastly, the main feature of CCA is that it involves listing all cost categories and outcome measures in a disaggregate format. A key difference between CCA and CBA is that although CCA reports costs and outcomes, it does not compare costs to outcomes.

Economic evaluations in healthcare typically focuses on maximising health, while public health interventions aim to prevent ill health, promote healthy behaviour and reduce inequalities. More specifically, public health interventions aim not only to improve health but also to improve a person's social circumstances (e.g. education, housing, pollution) (Marmot and Allen, 2014). Public health's focus on the social determinants of health means there may be several outcome measures which can be reported. This implies CBA or CCA may be the most appropriate approaches for analysing public health interventions. Nevertheless, there is a lack of consensus on how analysts should address the methodological challenges associated with CBA, such as how to assign monetary values to social outcomes and how to account for the unfair distribution of income in society (Donaldson et al., 2002). Similarly, for CCA there is a lack of consensus on how a decision-makers can implement the wide range results of CCA provides without the decision-maker's bias influencing which results they choose to focus on.

To date, there is no consensus in the literature for which economic evaluation methods should be used for public health economics. It is argued that due to the complex nature of public health interventions in terms of the multiple sectors, outcomes and components, there is no 'one size that fits all' (Edwards and McIntosh, 2019: 342). Some analysts choose evaluative frameworks that stem from the finance and accounting literature, including return on investment (ROI) and social return on investment (SROI). The appeal of ROI and SROI is that unlike CBA they provide a practical framework that is relatively straightforward for analysts with limited training in health economics to apply in practice. The limitations of ROI and SROI is that they lack the theoretical underpinning of welfare economics and ignore the methodological challenges which health economists have debated for over forty years (Fujiwara, 2015).

1.5. Welfarism and extra-welfarism

1.5.1. Welfare economics

Boadway and Bruce (1984) refer to welfare economics as the systematic study of methods (e.g. frameworks) used to order and rank society's preferences for any set of arrangements. A set of arrangements can include different states of the world and alternative courses of action (e.g. allocation of resources). The social ordering and ranking characteristic of welfare economics means the methods that underpin it are normative. This implies individual's value judgements are

needed to order and rank preferences for alternative courses of actions. This allows alternative courses of action to be meaningfully compared against each other. Furthermore, this meaningful comparison enables the analyst to make statements about whether one course of action is better, worse or equally as good as the alternative courses of actions (Boadway and Bruce, 1984).

1.5.2. Welfarism and extra-welfarism

The most dominant framework for comparing and ranking different states of the world (alternative courses of action) as better, worse or equal is called welfarist economics. Welfarist economics is underpinned by utility-related principles, where utility is synonymous for 'happiness' or 'satisfaction' (Morris et al., 2012). More specifically, welfarism is underpinned by the following principles: (1) the individual makes rationale decisions in order to maximise their welfare; (2) the individual is the best person to decide on how to improve their own utility (happiness), not others (e.g. government); (3) utility comes from the outcome of a choice made to maximise welfare as opposed to the process of making the decision itself; and (4) judging the value of something can only be judged using utility-based outcomes meaning all non-utility outcomes are irrelevant (Culyer et al., 2012). As welfarist methods are underpinned by individualism and the idea that only the individual can state what maximises their utility, welfarist economics is difficult to apply to health care and public health. This is because in many countries, it is the health professional who makes rational choices and decisions on behalf of the individual (Morris et al. 2012) as health and public health are highly specialised areas of study. Furthermore, in many countries, health and public health services (goods) are publicly funded which means decision-makers need to make comparisons between individuals in order to allocate resources across the public.

Consequently, welfarism has been described as a restrictive evaluative framework compared to other frameworks such as extra-welfarism (Culyer et al., 2012). For example, extra-welfarism permits comparisons to be made between individuals as well as permitting utility to be judged and specified by others and not only the individual who is experiencing the outcome. Furthermore, extra-welfarism permits the analysis of not just utility, but also a broader range of non-utility outcomes including characteristics such as health status, capabilities, and other issues of concern that go beyond an individual's utility (Culyer et al., 2012). For these reasons, extra-welfarism is described as a pragmatic framework since in practice it can be applied to a range of public policy challenges (Brouwer and Koopmanschap, 2000).

*1.5.*3*. Cost-utility analysis*

The pragmatic approach of extra-welfarism explains why it has been most commonly drawn on to address resource allocation problems within the field of health economics. In health economics, the most dominant extra-welfarist framework used is a species of cost-effectiveness analysis (CEA), called cost-utility analysis (CUA). It is important to note that in CUA, utility represents a measure of health characteristic. CUA measures health-related quality of life (HRQoL) not overall quality of life. The definition of utility in CUA therefore differs to the welfarist's definition of utility which refers to an individual's satisfaction. Analysts point out that use of the term utility in CUA is unhelpful as it is misleading (Culyer et al., 2012). A key reason for the widespread adoption of the CUA in health economics, may relate to the way measuring health characteristics (utility) enables comparisons

between individuals to be made. Comparisons are important for health professionals and public policymakers since in most health economies, health professionals and policymakers make decisions on behalf of others due to the high level of knowledge and expertise required to make healthcare decisions.

Health characteristics in CUA are measured using a multi-dimensional preference-based outcome measure. These measures captured include generic HRQoL measurements (e.g. EQ-5D) and disease specific quality of life measures (e.g. EORTC-8D). The EQ-5D which measures generic HRQoL, is the most commonly used instrument for measuring health characteristics (Rabin et al., 2011). The preference-based element of the measure relates to the way in which the different characteristics of health can be assigned utility (preference) weights. These weights are cardinal numbers and represent the relative importance of each dimension from the quality of life tool. Specifying a set of utility weights is a normative process as it requires subjective value judgements to be made by a general or disease-specific population (Dolan, 1997). Application of the set of utility weights to an individual's EQ-5D score is advantageous compared to non-preference based health measures, as the utility weights enables the analyst to infer whether a person's health has improved or deteriorated and by how much (Dolan, 1997). On the contrary, the disadvantage of CUA is that the term utility within the CUA framework does not fully reflect the individual's preferences and does not include all dimensions of health only those incorporated within the simple measurement tool. CUA is therefore seen as a pragmatic approach as it is practical (Brouwer and Koopmanschap, 2000). Nonetheless, CUA does not address all theoretical considerations.

1.5.4. Cost-benefit analysis

Although CUA uses a generic outcome measure which enables comparisons to be made across all disease areas, CUA has been criticised for not capturing the broader non-health outcomes which are important to individuals, families, communities and society (Edwards et al., 2013a). A framework is therefore required to specify what health and non-health outcomes are important, and how they should be measured and valued relative to each other. This is the role CBA seeks to play. The CBA framework assigns a monetary value to all outcomes so as outcomes can be aggregated into a single monetary value and compared with all aggregated costs. If the monetary value of the benefits are greater than the costs, then this can be interpreted as an efficient allocation of resources in society. CBA is often used by Governments in Impact Assessments for new policies.

Although CBA is deemed theoretically superior than CUA, the approach is challenging to apply in health economic practice (Drummond et al., 2015a). In traditional economics, the market makes it possible to use a CBA framework to evaluate services. The market can be analysed to reveal the value individuals place on consumed goods. In health economies, the consumption of health (where health can be seen as a good) is distorted since individuals who consume health (e.g. patients) do not typically purchase health directly through the market. This means market prices do not exist for health and social outcomes (Drummond et al., 2015a). As an alternative to market prices, the CBA framework can use alternative approaches to generate monetary estimates, such as willingness to pay (WTP) methods. WTP methods use hypothetical scenarios to capture what prices people claim they would be willing to pay for a particular health outcome. The challenge of

applying a CBA approach within public policy is that the hypothetical compensation tests are determined by an individual's pre-existing income and health, which is distributed unevenly across society (Donaldson et al., 2002). As a result, some argue that extra-welfarism approaches such as CUA may be more fair than welfarist approaches such as CBA (Brouwer and Koopmanschap, 2000). This is because CUA provides a way to explicitly assign equity weights (utilities) to health states so as decisions can be based on need rather than an individual's WTP. In the absence of consensus on what framework should be used for the evaluation public health trials, it is agreed that although CUA has limitations, QALYs are useful at illustrating that public health interventions are very cost-effective under NICE's willingness-to-pay per QALY threshold compared to most curative interventions (Owen et al., 2012).

1.6. Need for guidance on conducting economic evaluations

1.6.1. Insufficient methodological guidance for PA and SB

The methods used in economic evaluations of PA interventions vary substantially which contributes to the mixed results on whether specific PA interventions are good value for money (Vijay et al., 2016, Williams et al., 2012). Reasons for the variation in the approaches to economic evaluations include the fact that researchers come from different countries which support different methodological principles and practices (Torbica et al., 2017). Furthermore, the normative and pragmatic nature of economic evaluations, as well as the relative infancy of the methods, means researchers are required to make a number of methodological choices and assumptions which are normative and specific to the intervention, setting and population being evaluated. Examples of these choices and assumptions include: what effects and costs should be included in the analysis, how inflation should be accounted for (Crowley et al, 2014), how future costs and benefits should be discounted, and how uncertainty should be characterised (Weatherly et al. 2009).

In addition, four additional methodological challenges have been identified as being unique to the conduct of economic evaluations of public health interventions, which includes the conduct of PA and SB economic evaluations (Weatherly et al., 2009, Alayli-Goebbels et al., 2014). These four challenges are explored further in Chapter 2. Addressing these challenges requires adequate economic expertise and resources, which may explain why some researchers do not presently include a cost-effectiveness analysis in their evaluations. This may also explain why most clinical practice guidelines are primarily informed by the evidence from effectiveness evaluations since there is a lack of economic data being generated (van Mastrigt et al., 2016). Existing tools are available for assessing cost-effectiveness of PA interventions at the policy level, these include the MOVE (2.0) tool (Sport England, 2016) and HEAT tool (World Health Organisation, 2014). Nevertheless, the tools are restricted to the assessment of just two types of costs (immediate intervention operating costs and long-term disease costs). Furthermore, the tools require the user to have an aggregate unit cost for the intervention. The tools do not provide guidance on how to calculate the aggregate unit cost. Performing a microcosting exercise of new and complex interventions can be challenging as human resources (staff's time) make up a large proportion of the costs (Glick et al., 2014, Mogyorosy and Smith, 2005).

1.6.2. Multidisciplinary understanding of economic evaluation in PA and SB

There is widespread interest in economic data with the number of trials collecting economic data increasing (Ramsey et al., 2015). In the public sector, year on year the demand for cost data continues to grow (Curtis & Burns, 2018). Reasons for this include the fact that in a number of countries' public sector budgets are tighter and healthcare demand is increasing (Weatherly, Cookson, & Drummond, 2014). In the UK, it is reported that Public Health Directors are 'hungry' for economic evidence on the short-term economic impact of the preventative interventions they commission (Willmott, Womack, Hollingworth, & Campbell, 2016). Similarly, health and social care bodies such as the National Institute for Health and Care Excellence (NICE) in the UK request economic evidence to support their decision-making (NICE, 2014a). Furthermore, the MRC guidelines for the conduct of evaluations of complex health and public health interventions name assessment of cost-effectiveness as a key stage in the evaluation process (Craig et al., 2008). Despite calls for economic evaluation to be routine practice in the evaluation of public health (Wanless, 2004), the evidence is scarce compared to the evidence on clinical effectiveness. In part, this may be due to the international shortage of health economic expertise. That is to say, there is a need for health economic training in order to meet the demands of current public health challenges (Frew et al., 2018). One reason for the lack of economic evaluations in the field of PA and SB may be due to the lack of multidisciplinary working across research centres (Davis et al., 2014). Multidisciplinary working is important, as economic evaluations of PA and SB intervention cover a range of disciplines including: public health, exercise science, economics, policy and behaviour change (Davis et al., 2014).

1.6.3. PA and SB as a priority area

As public resources are scarce, evidence on the cost-effectiveness of PA and SB interventions is required to prevent politicians from disinvesting in highly cost-effective interventions. For example, in 2015, UK politicians disinvested £200 million in local public health budgets, which is forecast to cost the health sector £1 billion in the long-run (Allen, 2015). If economic evidence is not provided to demonstrate how PA and SB interventions can lead to large health gains and cost savings in the future, then politicians cannot be held accountable for disinvesting in PA and SB.

PA and SB interventions are a priority area to gather economic evidence on since the latest UK guidelines of PA highlight that there is no minimum amount of PA required to achieve some health benefits (Gibson-Moore, 2019). If a range of PA/SB interventions are increasing PA by even just a small amount, then there is a need for economic evaluations to help decision-makers understand whether the increased amount of PA can be regarded as good value for money compared to the amount of PA achieved through alternative interventions.

Economic evaluations of PA and SB interventions targeted at individuals who are highly sedentary is a particularly important area of study compared to other public health challenges. This is because these interventions have the potential to be highly cost-effective due to the curvilinear relationship between PA and health benefits (Warburton and Bredin, 2016). The curvilinear relationship shows that the less active an individual is prior to a PA intervention, the greater the health gains are for this individual. Furthermore, PA and SB intervention targeting individual's who

are the least active in society align with the principle of proportional universality, the belief that greater efforts and resources should be allocated towards those who are in most need and who face barriers to participating in PA (World Health Organisation, 2018).

1.6.4. The uniqueness of economic evaluation in PA and SB

PA and SB are particularly challenging field of study in economic evaluation as PA and SB interventions are typically delivered in non-healthcare settings. More specifically, the four domains where PA and SB interventions are delivered, include: at work (e.g. walking meetings), at home (e.g. housework), through transportation (e.g. walking to work) and through leisure (e.g. dance classes at a leisure centre) (Strath Scott et al., 2013). This requires the researcher conducting the economic evaluation to have an understanding on the setting and organisation in which the intervention is implemented (e.g. workplace). Furthermore, PA and SB are both multi-dimension behaviours. For instance a PA/SB intervention may attempt to modify an individual's frequency, intensity, time and/or type of PA (Barisic et al., 2011). Modifying these dimensions is likely to modify the amount of resources (e.g. time, materials, payments) required to deliver an intervention which may impact on the overall cost of the intervention.

Overall, the multiple domains and dimensions of PA and SB requires that researchers carrying out economic evaluations in this area are aware of existing practical tools which can help them capture and understand the complex nature of the PA and SB intervention. Furthermore, studying the complex nature of PA and SB interventions requires a framework which encourages a multidisciplinary approach, since PA and SB cut across multiple disciplines and research fields, including: public health, health economics, anthropology, physiology, trial methodology, behaviour change. Importantly, a framework needs to describe an approach which can be applied in practice and explain why the approach is appropriate. Transparent reporting on what and why a method is recommended needs to be a key feature of a framework for PA and SB, so as those using the framework can critically appraise the approach outlined. This PhD project aims to develop a framework to support the conduct of economic evaluations for individual-level PA and SB interventions.

As well as a framework which draws on knowledge and methods from multiple disciplines, the practicability of the framework will be assessed by piloting an initial framework to two case studies. These two case studies are: (1) an individual-level PA intervention delivered in a leisure centre (the leisure domain) which targets inactive individuals who have pre-existing health conditions; and (2) an individual-level SB intervention delivered in a contact centre (workplace domain) which targets SB amongst office workers. Targeting individuals with pre-existing health conditions is important as they are twice as likely to not be active enough to benefit their health than the general population (Public Health England, 2019). Similarly, office workers are also an important population to target as research shows that office workers reportedly spend upto 70-85% of their working day sitting (Healy et al., 2013, Clemes et al., 2014, Morris et al., 2019). As discussed, due to the curvilinear relationship between PA and health benefits, individuals who are less active prior to a PA intervention are more likely to achieve the greatest health benefits from an PA or SB intervention (Warburton and Bredin, 2016). As these two case studies target two different population groups

and are set in two different settings (domains), this provides an important opportunity to generate empirical evidence on the applicability of my framework.

1.7. Epistemological background

1.7.1. Overview of existing skills and knowledge

Throughout this PhD project I acknowledge that my understanding on how knowledge can be generated and interpreted has been influenced by my previous training and research experience. Prior to starting this PhD I had discipline-specific knowledge in anthropology and public health, having studied these two subjects for my undergraduate and postgraduate master degrees, respectively. Furthermore, prior to undertaking this PhD project, I was employed for almost three years as a Research Assistant in Public Health at Liverpool John Moores University. Through this employment, I gained a substantial amount of practical experience in the conduct and analysis of structured qualitative methods (e.g. focus groups and interviews) and the application of social return on investment (SROI) methodology.

This present PhD project in health economics presented me with the opportunity to train in a new field. It also presented me with the opportunity for me to build on my existing anthropology and public health skills and knowledge and develop as a true multidisciplinary researcher. Particular skills and knowledge I possessed before starting this PhD included: (1) an awareness of our need to understand the complexity associated with changing an individual's lifestyle and cultural norms, through my study of anthropology; (2) the burden of noncommunicable disease and inequity on society, through my study of and employment in public health; and (3) the importance of capturing the wider social, economic and environmental benefits of services and involving stakeholders in the data capture process through my SROI work. These existing skills and knowledge have been particularly complementary to this PhD as the project focuses on complex individual-level PA and SB interventions. In the subsequence sections, I provide an introduction to the methods and paradigms that underpin the disciplines of anthropology and public health with the intention providing insight into the methods and beliefs I had when I began my PhD journey.

1.7.2. Studying complexity

Anthropology is the study of society and culture. Society refers to the way humans organise themselves in a meaningful way, such as their patterns of interactions and power relationships. Culture refers to the behaviour, beliefs and values adopted by members of society. Societies and culture are complex (Hylland-Eriksen, 2001). A key aim of anthropology is to describe how complex the real world is, rather than trying to simplify it (Hylland-Eriksen, 2001). More specifically, anthropology aims to document the complex details of everyday life in which interventions, trials and policies are implemented, in order to help us understand what is appropriate and why certain results have been realised (Lambert and McKevitt, 2002). That is to stay, my anthropological background led me to believe it would be valuable to document the complexity of the everyday context in which my PhD framework is developed, through the use of anthropological methods which are typically informal and less structured approaches.

1.7.3. Naturalism and normative statements

A key contribution of the discipline of anthropology, is the ethnographic methods which derive from the discipline, namely participant observation. In participant observation the researcher participates in the research activities so as they can document what actually happens. Participant observation helps distinguish between normative statements (i.e. what interviewees and focus group members say should be done) and actual practice (what was actually done) through the use of informal and less structured research methods such as observations, informal conversations and notes made in the field (Lambert et al. 2002). Lambert and McKevitt (2002) argue for greater use of informal data collection rather than relying solely on formal and structured qualitative methods. They explain that structured qualitative data, such as an interviewee's normative statement, cannot be taken at face value. Furthermore, they claim that methods such as one-off interviews do not typically provide the broader context for an interviewee's answer. The authors argue that there may be a difference between what the interviewee says and what actually happened, and that we can only know this by observing the action for ourselves. This is where methods such as participant observation have a key role in helping us understand this knowledge gap between what people say they do or plan to do, and what people actually do. That is to say, anthropology believes it is important to observe actions and events in their natural context through ethnographic methods (Hylland-Eriksen, 2001). Some refer to this belief as the 'naturalism' paradigm. The naturalism paradigm suggests that people's actions, behaviour and values documented in their natural context without intervention (e.g. an RCT) is a valid source of knowledge. Some even argue that all phenomena should be studied in its natural state rather than in an artificial state (Atkinson and Hammersley, 2007).

1.7.4. Positivism and evidence-based medicine

The positivist research paradigm involves the testing of theories in artificial settings through experimental trials or through statistical control. This paradigm contradicts the belief of the naturalism paradigm which suggests actions should be observed in their natural settings. RCT methods traditionally belong to the postitivist paradigm, which is the belief that the best way to generate knowledge is through the use of standardised quantitative methods which can be replicated. In the 20th century, quantitative statistical methods developed rapidly which increased the popularity of positivist methods (Atkinson and Hammersley, 2007). In the second half of the 20th century, positivist methods continued to grow in popularity. There was a call in the discipline of medicine, for all evidence to be derived from positivist methods such as RCTs and systematic reviews which involve the testing of hypotheses through physical control, and standardised and replicable methods (Cochrane, 1972). Advocates of this new belief in what constituted 'reliable knowledge' claimed a new paradigm in medical practice was emerging and called this paradigm 'evidenced-based medicine' (Guyatt et al., 1992). Prior to conducting this PhD I had not been involved in any clinical trials. That is to say, the Co-PARs and SLaMM trials provided me with the opportunity to gain practical experience in how experiments are set up and delivered in practice. Through my employment as a Research Assistant, I had been involved in a number of systematic reviews in the field of public health. I therefore understood the value of generating evidence from more positivist research methods such as systematic reviews which gather data from a large sample of studies through a structured and replicable process.

1.7.5. Multidisciplinary working and pragmatism

My masters degree and employment in public health, introduced me to the idea that multidisciplinary and pragmatic research methods can improve our understanding of public health and health inequity. A recurring piece of literature discussed throughout my masters degree was the Marmot Review (2010). This review furthered my interest in inequity and made me aware of the relationship between the four major lifestyle diseases in the UK (one of which included physical inactivity) and socioeconomic status. In relation to my existing multidisciplinary experience, for masters degree I studied a broad range of disciplines including modules in statistics, epidemiology, public policy, health economics, sociology and psychology. This made me aware of how the data which is generated to improve our understanding of a public health challenge such as physical inactivity, can depend on the type of disciplines and research methods used to evaluate the public health problem. For example, through my masters I learnt that epidemiological methods traditionally study trends across time, places and people, and that psychology methods traditionally test hypotheses in artificial settings. This helped me recognise that these two different approaches will lead to different types of knowledge being generated. Moreover, through my public health employment I learnt more about mixed methods research and how it has an important role in addressing public health challenges. More specifically, I gained practical experience of triangulating qualitative and quantitative data and seeing how the two different methods can be complementary. As I was aware of the strengths of a multidisciplinary approach to addressing public health challenges, this PhD project aligns with the pragmatic paradigm. This means that the research methods I select to use throughout this thesis are be those which I believed are most appropriate for the research question and the context in which they are being applied (Johnson and Onwuegbuzie, 2004).

1.7.6. Reporting the development of the framework

John and Smith (2017) argue that one of the key contributions of anthropology is its generation of descriptive ethnographic data which documents and provide explanations on how and why actions occur. Throughout this thesis I have reported ethnographic data in the form of reflection boxes. The reflection boxes provide insight into how and why I have developed my systematic framework in a particular way. Reflective content has been be based on the notes I make in the ‘field’ during my PhD project. The ‘field’ in this context will refers to the everyday observations and interactions I had. These interactions include the informal meetings I had with stakeholders associated with my PhD, namely the researcher team linked to my PhD project (e.g. my supervisory team and the other postgraduate researchers), the health economists from Deakin University’s Health Economics group (Australia), and the staff from the local authority, leisure centres and workplace company in which my two trials were set.

1.8. Aim of PhD

The overarching aim of the PhD was to develop a multidisciplinary and pragmatic framework to support researchers from multiple disciplines to conduct economic evaluations of individual-level PA and SB interventions. These multiple disciplines include those involved in addressing the challenge of physical inactivity, namely: physiologists, psychologists, public health professionals

and trialists. Whilst the framework will be intended for researchers with limited or no specialist training in health economics, the framework may also be of use to health economists seeking a standardised approach to economic evaluation of individual-level PA and SB interventions. The three key objectives in order to achieve this aim were to:

1) Conduct a systematic review of existing economic evaluations of individual-level PA and SB interventions in order to explore how analysts have addressed four key methodological challenges, which are regarded as being unique to the conduct of economic evaluations of public health interventions (Study 1, Chapter 2).
2) Develop an initial framework which draws on good-quality methods identified from the systematic review (Chapter 3)
3) Pilot the initial framework concurrently in two individual-level trials, with the aim of:
    i. Reflecting on the applicability of the framework to a PA trial (Study 2, Chapter 4) and SB trial (Study 3, Chapter 5)
    ii. Providing evidence on the cost-effectiveness of a PA on referral scheme intervention (Study 2, Chapter 4) and a workplace SB intervention (Study 3, Chapter 5)

Chapter 6 synthesises the empirical findings from the studies 1-3 in order to recommend a refined version of the framework, make recommendations for future research, and consider implications for policy and practice. Figure 1 shows how each objective feeds into the subsequent studies and chapters.

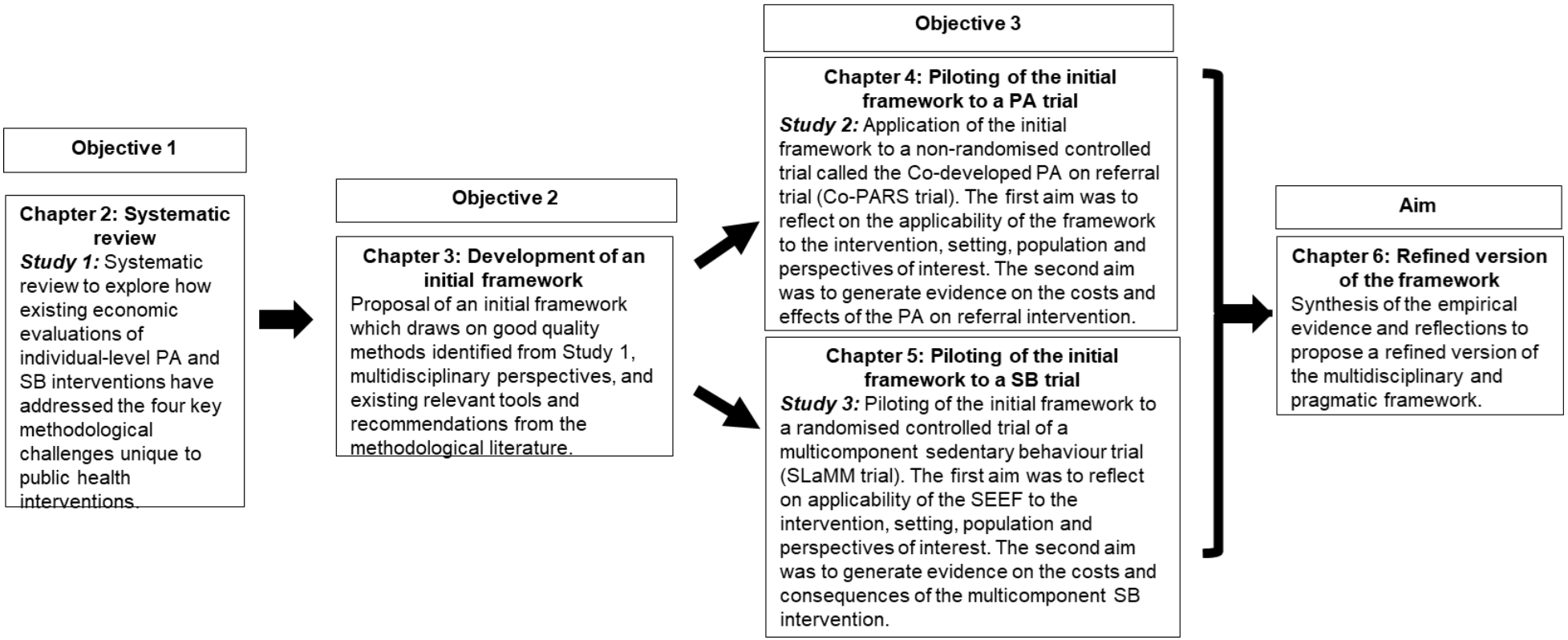


Figure 1. Structure of PhD

Chapter 2: Systematic review of the methods used in economic evaluations of individual-level PA and SB interventions

2.1. Introduction

Despite recommendations for economic evaluations to become routine within public health interventions (Kelly et al., 2005) cost-effectiveness information on PA and SB interventions remains scarce (Abu-Omar et al., 2017). One reason for this lack of analysis may be due to the lack of guidance and multidisciplinary efforts to inform analysts on how to conduct economic evaluations in the field of public health (Davis et al., 2014). Economic evaluations of public health interventions are subject to four key methodological challenges identified and described in former reviews (Hill et al., 2017, Alayli-Goebbels et al., 2014, Weatherly et al., 2009) as: attribution of effects; measuring and valuing outcomes; identifying intersectoral costs and consequences; and incorporating equity.

The first review to explore the economics of public health was conducted by Weatherly et al. (2009). These authors were the first to specify and name the four key methodological challenges for public health economic evaluations. The authors identified the four challenges after reviewing five reviews which discussed the economics of public health. The inclusion criteria for the review by Weatherly et al. (2009) was broad as it included economic evaluations from 11 public health areas. That said, the authors only reviewed studies published between 2000-2005. In total, 154 NHS Economic Evaluation Database (NHS EED) abstracts were examined, of which 53 related to the field of obesity and PA. The study identified four studies (3%) which claimed to be CBA, but the review authors reported that after further examination these were not CBAs but three CCAs and one CUA. Overall, the review authors claimed they gained little insight on how to address the four methodological challenges for public health studies.

The review by Alauli-Goebbels et al. (2014) refers to methodological challenges identified by Weatherly et al (2009) and also assesses methodological quality of the studies. Furthermore, the review focuses on six key behaviour change areas: smoking, PA, dietary behaviour, drug use, alcohol use and sexual behaviour. The authors carried out their searches in 2009 and identified 142 eligible studies which had been published between 1981-2009. Seventeen of these studies assessed PA. The authors reported that an overarching finding from their review was that that the studies do not always report sufficient details around the methods and study design they used. They explain that this made it difficult to see how studies had handled the methodological challenges. In relation to PA, the authors reported identifying studies which commented on psychological wellbeing being a broader outcome of PA interventions, however they reported that analysts did not incorporate this outcome into their economic evaluations.

The study by Hill et al. (2017) reviewed 27 economic evaluations and priority-setting studies in the field of alcohol prevention, published between 2006-2016. The reviewers concluded that studies in the field of alcohol prevention are not addressing the methodological challenges unique to public health challenges. They found that most studies did not consider: long-term outcomes, wider perspectives or equity. They also reported a lack of CBA, CCA and priority setting studies. That said the reviewers did identify one CBA study, although they report that the authors of the CBA do not explain how they monetised the health benefits included in the analysis.

Overall, all three reviews explore the empirical evidence to see how the methodological challenges associated with public health are being addressed in practice. A key observation, is that all three reviews discuss how the studies do not provide sufficient detail on the methods they carried out. Although the study by Hill et al. (2017) looks at these challenges more recently, they focus on alcohol prevention. That is to say, there has been no review published on the methods used in PA economic evaluations since Alauli-Goebbels et al. (2014) conducted their searches in 2009. Since 2009 there has been several reporting guidelines published in the fields of trial methodology (Schulz et al., 2010), intervention design (Hoffmann et al., 2014), economic evaluation (Husereau et al., 2013) and equity (Welch et al., 2017). That is to say, if authors are now required to report more detail on the methods they have used, this data has the potential to improve our understanding on the four methodological challenges can be addressed in PA-related studies. The four methodological challenges are described in greater detail in the subsequence sections.

2.1.1. Challenge 1: Attribution of effects

Randomised controlled trials (RCTs) are the gold standard for evaluating the effectiveness of an intervention. RCTs alone are however insufficient to inform long-term investment decisions in health systems aiming to be sustainable. This is because conducting experimental studies such as RCTs over many years or decades is likely to be resource intensive from both the research funder and participant's perspective. Attrition from the trial and insufficient funding is inevitable. Yet, the greatest health outcomes and cost savings attributable to PA and SB interventions do not typically manifest until decades after an intervention has taken place. Due to this long pay-back time (Wanless, 2004), it is recommended economic evaluations link up trial-derived intermediate or surrogate outcomes with additional sources of evidence (e.g. observational studies) (Ramsey et al., 2015).

2.1.2. Challenge 2: Measuring and valuing outcomes

Previous PA studies have used different outcomes, or have classified the same type of outcomes in different ways, which makes it challenging to meaningfully use cost-effectiveness results and compare interventions (Abu-Omar et al., 2017). This is likely to be because PA and SB interventions are associated with a broad range of outcomes, many of which are not captured in evaluations that conduct just one type of valuation analysis. Furthermore, many broader important and relevant outcomes such as improved wellbeing or someone's ability to return to work are difficult to assign a monetary value, as they do not have a market price (Weatherly et al., 2014).

2.1.3. Challenge 3: Identifying intersectoral costs and consequences

Many PA and SB interventions take place outside of the healthcare setting, necessitating a time and equipment commitment from intervention participants and providers (which has an opportunity cost). Moreover, PA and SB interventions are complex, impacting on multiple sectors simultaneously (Dahlgren and Whitehead, 1991). Therefore, it is important to consider the impact of these interventions on other stakeholders including public sector agencies beyond the health sector, private individuals and the voluntary sector (Weatherly et al., 2014, Weatherly et al., 2009). Yet, as there is no universal definition for each perspective type, the costs and consequences

deemed relevant for inclusion in the analysis is primarily analyst-dependent (Husereau et al., 2013).

2.1.4. Challenge 4: Incorporating equity

A key objective in public health is to reduce inequity, meaning inequalities that are avoidable, but have not yet been avoided and are therefore unfair (Marmot and Allen, 2014). By contrast, a key objective in economic evaluation is to maximise efficiency across the whole population (Weatherly et al., 2014). If authors fail to acknowledge equity by not adapting their existing economic analysis approach, it is not transparent which socio-economic group have gained or lost out due to a resource allocation decision. Until the recent publication by Cookson et al. (Cookson et al., 2017) recommendations on how to incorporate equity have been limited within international and national guidelines for economic evaluation (Sanders et al., 2016, Ramsey et al., 2015, NICE, 2014a, Husereau et al., 2013). Approaches for incorporating equity into the analysis described by Cookson et al. (2017) include: equity impact analysis, equity constraint analysis and equity weighting analysis.

2.1.5. Aim

In an attempt to learn how the four challenges outlined above have been addressed in practice, this systematic review aims to provide an overview of the methods used in economic evaluations of PA and SB interventions since 2009. Alayli-Goebbels et al. (2014) and Weatherly et al. (2009) reviewed the methods reported in economic evaluations of a range of public health areas including 17 and 26 PA economic evaluations published up to 2005 and 2009, respectively, but the reviews found little insight from the empirical evidence. Economic evaluation is a rapidly developing field especially with the growth of decision-analytic modelling and the economic evaluation reporting standards (Drummond et al., 2015b, Ramsey et al., 2015). Accordingly, there is a strong rationale to provide an update on methods carried out since 2009.

2.2. Methods

2.2.1. Information sources and search strategy

A comprehensive search took place across six electronic databases that host reports from the medical and economic field (Medline via Ovid; SPORTSDiscus, EconLit and PsycINFO via EBSCO*Host*; NHS EED and HTA via the Cochrane Library). The database NHS EED stores records up to April 2015, thus searches in this database went up to 2015 only. Additional, supplementary searching was performed: key websites were searched for studies that included specific free text terms: ‘PA’, ‘SB’, ‘economic’ and ‘cost’; reference lists of two relevant systematic reviews (Gc et al., 2016, Wu et al., 2011) were hand searched; and protocols that met the majority of the eligibility criteria were used to search for completed studies via online searching and contacting the authors. An example of the full electronic search strategy for Medline is provided in Appendix A.1. This search was replicated for all databases, with amendments made as appropriate to align terms with individual database index terms.

2.2.2. Study selection

The protocol for this review can be retrieved from the PROSPERO database for registered systematic reviews (registration number CRD42017074382). Full economic evaluations of

interventions targeting individuals aged 16 years or over, who are defined as being physically inactive or sedentary, were eligible for inclusion in the review. Population level interventions were excluded as well as protocols. Eligible studies needed to capture PA and SB at two or more time points to observe if a change in behaviour has occurred. Comparators could be any alternative intervention including no intervention. Interventions and comparators targeting multiple behaviours such as PA and diet were excluded unless the multiple behaviours were PA and SB. Both trial and model based economic evaluations were eligible. Letters to editors and conference briefings were excluded. Both published and unpublished 'grey' literature were included. Abstracts where the full text could not be retrieved were excluded. Only English language studies were included due to the restricted language skills of the reviewers available. Eligibility criteria was applied during both screening phases. The present systematic review identifies and discusses studies published from January 2009 to March 2017. In addition, a rapid systematic scoping search was performed in Medline to understand whether new studies had been published in this area from March 2017 to January 2019. Details on methods of the scoping search are not discussed below, rather they are presented in Appendix A.2.

2.2.3. Screening

During the title and abstract screening phase two reviewers (first author, seventh author) screened 10% (n=612/ 6,123) of the studies and there was a disagreement rate of 2.94% (n=18). Reviewers discussed the disagreements and resolved them without the need to seek the expertise of a third reviewer. Reviewer one (first author) went on to screen the rest of the studies, informed by the disagreement discussions. Similarly, during the full text screening phase reviewer two (seventh author) screened 10% (n=15/ 153) of the studies. There was disagreement for 33.33% (n=5) of the studies. The reviewers discussed the disagreements and again a consensus was met without the need for a third reviewer. Figure 2 shows an overview of the study selection process.

2.2.4. Data extraction

A data extraction form was developed based on the items featured on the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (Husereau et al., 2013). The form was piloted independently by two reviewers (first author, seventh author) on two (10%) randomly selected studies. Following discussions the form was shortened, and items relevant to the four methodological challenges, and key study characteristics were retained. Following the piloting stage, the first reviewer extracted data for the remaining studies. A template of the final data extraction form is provided in Appendix A.3. It was not necessary to request additional information from the study authors.

2.2.5. Quality assessment

Drummond's 10-item checklist was selected as it is one of the most widely used quality assessment tools (Drummond et al., 2015b). A component approach was used when applying the checklist in Appendix A.4. This approach is advocated in the PRISMA statement and entails assessing each item individually rather than generating a summary score (Liberati et al., 2009). Two reviewers (first author, seventh author) independently conducted the quality assessment for 10% (n=2/ 15) of the included studies. Disagreement was limited to item 6 (Item 6: Were costs and

consequences valued credibly?) on the checklist, examples in Drummond et al. (2015b) were consulted to overcome these disagreements. Practical application of item 10 (Item 10: Did the presentation and discussion of study results include all issues of concern to the users?) was challenging due to the limited guidance, thus findings from this question are less informative. Alayli-Goebbels et al. (2014) also experienced this barrier in an earlier version of the checklist.

2.2.6. Method of analysis

The published narrative synthesis framework by Popay et al. (2006) guided the analysis to ensure a transparent and systematic approach was performed. The narrative synthesis in this review goes beyond describing how authors have addressed each of the four challenges by attempting to explain why specific approaches have been chosen. The analysis was an iterative process. A priori analysis involved tabulating the data and producing bar charts on key study characteristics: study design, time horizon, valuation technique, study perspective and explicit/ implicit equity analysis. The same study characteristics were focused on in the two former methodological reviews (Weatherly et al., 2009, Alayli-Goebbels et al., 2014). The wider literature also indicated that the following contextual factors were important to review when understanding an analyst's approach: intervention setting, country and year of publication. Additional ad hoc analyses were performed where trends became apparent. Lastly, the strength of the narrative synthesis and the conclusions derived from it were considered by reflecting on the quantity of studies and results of the quality assessment.

2.3. Results

A total of 15 economic evaluations (17 publications) were included in the review (Figure 2). Searching across Medline, SPORTSDiscus, EconLit, PsychINFO, NHS EED and HTA databases retrieved 7,063 records. Supplementary searching retrieved six additional records including: two records from hand searching on key websites, two from the reference list of a systematic review (Gc et al., 2016), and a further two from searching for the completed studies of two protocols (Kolt et al., 2009, de Vries et al., 2013) in Appendix A.5. After removing duplicates 6,129 records remained of which a further 5,907 records were removed as title and abstracts did not meet the eligibility criteria. During the full text screening, 159 citations were examined in further detail, of which 142 studies were excluded. Reasons are outlined in Figure 2.

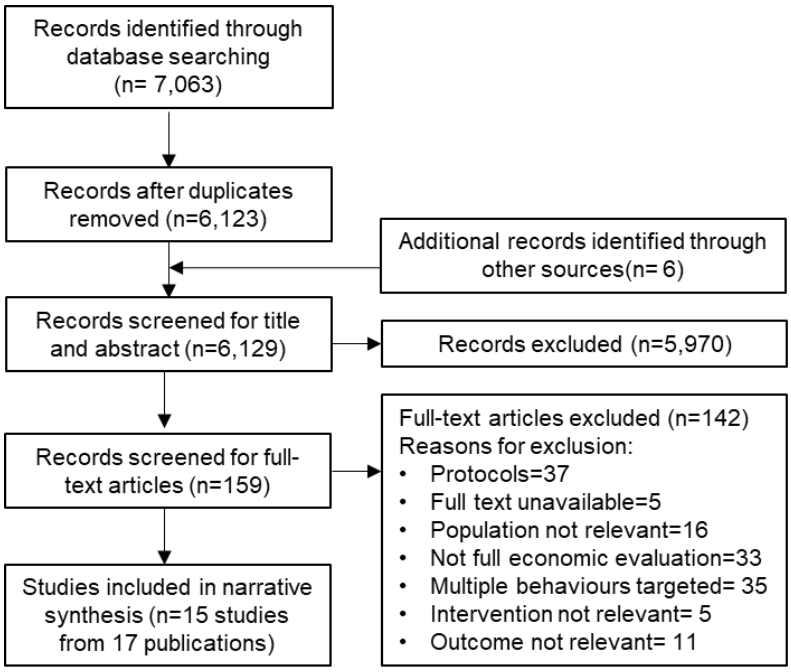


Figure 2. PRISMA flow diagram representing study selection process

2.3.1. Study characteristics

Of the 15 studies, ten were single trial-based economic evaluations and five were model-based; no studies were single trials that had extrapolated or modelled their results. Table 1 provides an overview of study characteristics for the trial- and model-based studies respectively. Studies are arranged by country followed by year of publication. Interventions were set in primary care, community and the home, and setting did not appear to be related to intervention type or country. As shown in Table 1, no studies targeted SB as an independent risk factor from PA. The range of interventions was limited to the following types: PA programme/ on prescription in primary care (n=9); brief advice in primary care (n=2); home-based informational advice (n=1); PA in a physical therapy setting (n=1); and fall prevention programme in both primary care and the home (n=1). The remaining study compared strategies for recruiting to PA interventions in primary care. The overall range of adult-based interventions matches the narrow range identified in a recent review of reviews focussing on the economic results of PA interventions (Abu-Omar et al., 2017). Studies came from four high-income countries. More than half (n=8) of the 15 studies came from the UK, with the remaining coming from New Zealand (n=3), the USA (n=2), and the Netherlands (n=2) (Table 1).

2.3.2. Quality assessment

Overall, studies performed well against Drummond's 10-item quality assessment checklist (Drummond et al., 2015b) in Appendix A.4. Nevertheless, six studies scored 'No' on at least one item: two studies did not state their perspective (item 1); three studies did not include all costs and consequences relevant to their stated perspective (item 4); one study did not discount its costs and consequences (item 7); and one study did not report their price source (item 6). Interpretation on whether item 4 was met by any of the ten trial-based economic evaluations who captured costs and outcomes at two years or less is up for debate. It could be argued that not all important and relevant costs and consequences can be identified for studies, which do not take a systems

approach (e.g. if they do not consider the impact on the wider system in which an intervention is being implemented nor capture the long-term impact) (Rutter et al., 2017, Squires et al., 2016). In order to align with other reviews which have used Drummond's checklist, the quality assessment results for item 4 were based on the checklist's accompanying guidance (Drummond et al., 2015b). Costs and consequences identified, measured and valued are discussed in greater depth in the subsequent sections.

Table 1. Overview of economic evaluations.

Trial-based economic evaluations								
Study & Year of publication	Stated perspective	Country	Population targeted	Sample size	Intervention	Comparator	Setting	Valuation technique
Iliffe et al. 2014	Health sector	UK	Inactive ≥65 years old who had fallen less than times in the previous 12 months	100	Falls Management Exercise Programme (Weekly group exercise class & 2 home-based exercise sessions)	Usual care (no intervention); Otago Exercise Programme	Primary care & community (as Home-based)	CEA
Edwards et al. 2013; Murphy et al. 2012	Multi-agency public sector	UK	Sedentary, and over 16 years, with risk factors for coronary heart disease, or mild to moderate anxiety, depression or stress.	798	ERS (primary care)	Information leaflet only	Primary care	CUA
Boehler et al. 2011	Health sector	UK	Inactive adults, 16 to 74 years old	46	Opportunistic recruitment strategy for PA interventions	Disease register strategy; Hypothetical no intervention strategy	Primary care	CEA

Shaw et al. 2011	Not reported	UK	Inactive, adults (age not defined)	79	Individualised walking programme: a pedometer and a 30-min consultation	Individualised walking programme: a pedometer, but and 5 min brief advice	Primary care	CEA
Larsen et al. 2015	Payer	USA	Inactive Latina women, 18-65 years old	266	Home print-based mail-delivered MVPA intervention linguistically and culturally adapted for Latinas	Wellness contact (information on health topics excluding MVPA)	Home-based	CEA
Young et al. 2012	Societal	USA	Women, following coronary artery bypass surgery	40	Symptom management intervention delivered by telehealth device to improve the PA level	Usual care, 2 week follow up call by the primary providers and cardiac specialists	Community	CEA
de Vries et al. 2016	Societal	Netherlands	Sedentary adults (or at risk of losing active lifestyle in near future) with mobility problems, ≥70 years old	130	Patient-centred physical therapy	Usual care for physical therapy, less patient-centred	Physical therapy setting	CUA

Maddison et al. 2015	Not reported	New Zealand	≥18 years old with diagnosis of IHD within previous 3- 24 months.	171	Exercise prescription and behavioural support by mobile phone text messages and internet	Usual care (participation in usual Cardiac Rehabilitation e.g. education session and psychological support)	Home-based	CEA; CUA
Leung et al. 2012	Public health system and participant	New Zealand	Inactive adults, ≥65 years old	330	Pedometer-based prescription, focus was on step-related goals	Green prescription, focus was on PA time-related goals	Community	CEA; CUA
Elley et al. 2011	Societal	New Zealand	Inactive, 40- 74 years old	974	Enhanced green prescription, 10 min of brief advice and a written exercise prescription with telephone support at 9 months and 30min face-to-face support at 6 months.	Usual care from GP (not standard green prescription, usual care from GP not defined)	Primary care	CEA

Model-based economic evaluations								
Study & Year of publication	Stated perspective	Country	Population targeted	Model type & size of simulation cohort	Intervention	Comparator	Setting	Valuation technique
Campbell et al. 2015	Health Sector	UK	Sedentary adults, ≥50 years old	Markov model (100,000 simulation cohort)	ERS (primary care)	Usual care (refers to Pavey et al. 2011's definition)	Primary care	CUA
Anokye et al. 2012; Anokye et al. 2014	Health sector; Health sector and participant for CCA	UK	Inactive, ≥33 years old	Markov model (100,000 simulation cohort)	Brief Advice (primary care)	Usual care (no intervention)	Primary care	CUA (and CCA)
Anokye et al., 2011	Health sector	UK	Sedentary adults, 40-60 years old	Decision tree model (1,000 simulation cohort)	ERS (primary care)	Usual care (refers to Pavey et al. 2011's definition)	Primary care	CUA

Pavey et al. 2011	Health sector CUA; Partial-societal for CCA	UK	Sedentary adults, 40-60 years old	Decision tree model (1,000 simulation cohort)	ERS (leisure centre)	Usual care (no active ingredient-PA advice or leaflets)	Leisure-centre	CUA (and CCA)
Over et al. 2012	Health sector	Netherlands	Inactive, 20-65 year olds	Markov model (100,000 simulation cohort)	GP pedometer prescription, counselling combined with pedometer use	Usual care (no intervention)	Primary care	CUA

ERS: Exercise Referral Scheme; GP: General Practitioner; MVPA: Moderate-to-vigorous PA; CEA: cost-effectiveness analysis; CUA: Cost-utility analysis

2.3.3. Challenge 1: Attribution of effects

Two thirds (n=10) of the studies in this review, all trial-based, did not compare the costs and consequences of the comparator groups beyond the trial follow up period (Table 2). More specifically, one study compared costs and consequences over a two-year period (Elley et al., 2011), the remaining nine had a time horizon of 12-months or less. For six of these studies, authors referred to their short time horizon as a limitation of their study (de Vries et al., 2016, Leung et al., 2012, Boehler et al., 2011, Larsen et al., 2015, Edwards et al., 2013b, Shaw et al., 2011). For instance, it precluded the incorporation of any potential long-term healthcare savings (Larsen et al., 2015). Just one study suggested future modelling exercises could be used to address this challenge (Edwards et al., 2013b). Yet, for Shaw et al. (Shaw et al., 2011) a short-time horizon was justified as they reported there was insufficient data to extrapolate their results over the participants' lifetime.

By contrast, all five model-based studies extrapolated a pooled trial-derived effectiveness estimate over the rest of the participants' lifetime; bridging the gap between the short- and long-term evidence (Table 2). Nevertheless, the assumptions underpinning the model-based studies varied considerably. Two studies (Anokye et al., 2011, Pavey et al., 2011b) made large assumptions unsupported by evidence about the duration of the effect, assuming that any short-term change in PA observed in the trials 6-12 months after the intervention, would be long-lasting. Over et al. (2012) employed a different approach by extrapolating an effect estimate, observed at 18 weeks, over a 40-year time horizon (the life expectancy of the participants). The authors assumed that only 25% of the effect recorded at 18 weeks would remain over the 40-year time horizon; they too reported that their assumptions were unsupported by evidence. These findings demonstrate how studies will vary according to the assumptions made. It is therefore important that end-users of cost-effectiveness results check they agree with the assumptions that underpin the economic evaluation.

Assumptions underlying the two other model-based studies (Campbell et al., 2015a, Anokye et al., 2012) were supported by three robust cohort studies. Campbell et al. (2015a) replicated Anokye et al.'s (2012) approach. More specifically, they linked the short-term change in PA level observed in trial data, with Hu et al.'s (2007, 2003, 2005) cohort studies that followed a group of active and inactive individuals for a duration of at least 10 years to predict how their activity levels and risk of disease changed over time. Anokye et al. (2012) explain how their identification and use of the cohort studies has strengthened previous modelling attempts in the field of PA. Campbell et al.(2015a) reported this approach has enabled more conservative assumptions to be made around changing PA levels and disease development over time.

Table 2. Time horizon and types of outcomes compared to costs

Trial-based economic evaluations		
Study & Year of publication	Time Horizon (trial follow up)	Types of outcomes compared to costs per valuation technique
Larsen et al. 2015	Trial duration (12 months)	CEA: Cost per minute of increase in PA
Iliffe et al. 2014	Trial duration (12 months)	CEA: Cost per participant reaching or exceeding 150 minutes of moderate-to-vigorous PA per week
Young et al. 2012	Trial duration (3 months)	CEA: Cost per incremental change in daily estimated energy expenditure; CEA: Cost per the incremental change in minutes spent on moderate-to-vigorous activity
Elley et al. 2011	Trial duration (24 months; 12 months)	CEA: Cost per participant achieving 150 minutes of moderate intensity activity per week
Boehler et al. 2011	Trial duration (3 months)	CEA: Cost per participant achieving 150 minutes of moderate intensity activity per week

Shaw et al. 2011	Trial duration (12 months)	CEA: Cost per additional person achieving the target of a weekly increase of ≥ 15,000 steps.
Maddison et al. 2015	Trial duration (24 weeks / [6 months])	CEA: Cost per MET-hour of walking and leisure activity; CUA: Cost per short-term QALY gain
Leung et al. 2012	Trial duration (12 months)	CEA: Cost per 30 minutes of weekly leisure walking; CUA: Cost per short-term QALY gain
de Vries et al. 2016	Trial duration (6 months)	CUA: Cost per short-term QALY gain
Edwards et al. 2013; Murphy et al. 2012	Trial duration (12 months)	CUA: Cost per short-term QALY gain

Model-based economic evaluations		
Study & Year of publication	Time Horizon (trial follow up)	Types of outcomes compared to costs per valuation technique
Campbell et al. 2015	Lifetime	CUA: Cost per short-term QALY gain (mental health gain); Cost per QALYs associated with coronary heart disease, stroke, type 2 diabetes due to reduced risk for developing these health states
Anokye et al. 2012; Anokye et al. 2014	Lifetime	CUA: Cost per short-term QALY gain (mental health gain); Cost per QALYs associated with coronary heart disease, stroke, type 2 diabetes due to reduced risk for developing these health states CCA: Same outcomes outlined below for Pavey et al.'s (2011) CCA
Anokye et al., 2011	Lifetime	CUA: QALYs associated with coronary heart disease, stroke, type 2 diabetes due to reduced risk for developing these health states
Pavey et al. 2011	Lifetime	CUA: Cost per short-term QALY gain (mental health gain); Cost per QALYs associated with coronary heart disease, stroke, type 2 diabetes due to reduced risk for developing these health states CCA: Mental health (anxiety), Mental health (depression), Metabolic diabetes, Colon cancer, Breast cancer , Lung cancer, Hypertension (cardiovascular), Coronary Heart Disease, Stroke, Musculoskeletal (Osteroporosis), Musculoskeletal (Osteroarthritis), Lower back pain, Rhumatoid arthritis, Falls prevention, Absenteeism at work, Injury (disbenefit), Disability
Over et al. 2012	Lifetime	CUA: QALYs associated with myocardial infarction, stroke, diabetes, colorectal cancer, breast cancer due to reduced risk for developing these health states

RCT: randomised controlled trial; cRCT: cluster randomised controlled trial; CEA: Cost-effectiveness analysis; CUA: Cost-utility analysis; CCA: cost-consequence analysis; MET: Metabolic Equivalent of Task

2.3.4. Challenge 2: Measuring and valuing outcomes

No studies in this present review conducted a cost-benefit analysis (CBA), despite health economists (Drummond et al., 2015b) stating this approach is superior to cost-utility analysis (CUA) (Drummond et al., 2015b). Recent UK and US guidelines recommended that studies report a broad range of outcomes alongside their economic analyses, through the use of approaches such as CBA, cost-consequence analysis (CCA) or an impact inventory (Sanders et al., 2016, NICE, 2014a). Two studies (Anokye et al., 2012, Pavey et al., 2011b) included a CCA conducted alongside a CUA. A broad range of health outcomes were included in their CCA (Table 2) yet the only non-health outcome reported was absenteeism.

Two thirds (n=11) of the studies presented just one type of valuation technique, either a CUA (n=5) or cost-effectiveness analysis (CEA) (n=6) (Table 2). Table 2 demonstrates further how despite having the same aim to increase PA levels and same valuation technique, the way results are presented to the end-user are inconsistent. Young et al. (Young et al., 2012) performed two CEAs reporting on the 'cost per incremental change in daily estimated energy expenditure' and 'cost per incremental change in minutes spent on moderate-to-vigorous activity'. Three other studies (Iliffe et al., 2014b, Elley et al., 2011, Boehler et al., 2011) performed a different type of CEA reporting on 'cost per participant achieving 150 minutes of moderate PA per week'. The most common way to present the result of the valuation analysis was as 'cost per short-term quality-adjusted life year (QALY) gain'. Nevertheless, this was reported for just under half (n=7) of the economic evaluations: four trial-based (Maddison et al., 2015, Leung et al., 2012, Edwards et al., 2013b, de Vries et al., 2016) and three model-based (Campbell et al., 2015a, Anokye et al., 2012, Pavey et al., 2011b) studies. All model-based studies conceptualised the long-term gain in QALY in the same way, in terms of the QALYs gained due to not developing coronary heart disease, stroke or type 2 diabetes, or experiencing premature mortality. Over et al.'s (2012) analysis differed slightly, as they also included colorectal and breast cancer.

Rationale for the inclusion and exclusion of trial-derived QALYs varied considerably. Shaw et al. (Shaw et al., 2011) argued against the inclusion of trial-derived QALYs in their analysis, explaining it would be unnecessarily restrictive since evidence already shows that PA is associated with a reduction in NCD and premature mortality, which in turn is associated with a much greater gain in QALYs than trial-derived QALYs. Three model-based studies (Campbell et al., 2015a, Pavey et al., 2011b, Anokye et al., 2012) deemed it appropriate to incorporate both short-and long-term gain in QALYs. They conceptualised the short-term QALY gain as being a one-off gain in mental health, which they assumed would be achieved as a result of becoming physically active for at least 90 minutes per week. They assumed the one-off mental health benefit would last for just one year, which they claimed was a conservative assumption. Campbell et al. (2015a) reported that their cost-effectiveness result was highly sensitive to the inclusion and exclusion of the one-off gain in mental health.

2.3.5. Challenge 3: Identifying intersectoral costs and consequences

The most commonly reported perspective was the health sector perspective (n=7) (Table 1). Six of the eight studies from the UK were from this perspective. In 2014, the UK reference case was

updated to recommend the public sector perspective when conducting economic evaluations of public health interventions (NICE, 2014a). The multi-agency public sector perspective adopted by Edwards et al. (2013b) reflects the start of this paradigm shift. Two more recent UK studies (Campbell et al., 2015a, Iliffe et al., 2014b) did not adopt a public sector perspective. Despite studies being conducted from the same perspective, the type of costs identified as relevant varied within and across countries and intervention type. This weakness was identified through the quality assessment (Item 4 on Appendix A.4), as five studies (Young et al., 2012, de Vries et al., 2016, Boehler et al., 2011, Maddison et al., 2015, Shaw et al., 2011) did not relate their costs to a study perspective. More specifically, two studies did not report their perspective (Shaw et al., 2011, Maddison et al., 2015) and three included a narrower range of costs and consequences than would be expected for their stated perspective (Boehler et al., 2011, Young et al., 2012, de Vries et al., 2016). For example, two studies stated their study was from the societal perspective yet assessed only direct intervention costs and short-term healthcare savings (de Vries et al., 2016, Young et al., 2012), which were the same costs as studies which stated taking a health sector perspective (Table 1) Weatherly et al. (2009) also found that many studies included only a narrow range of costs within their stated study perspectives.

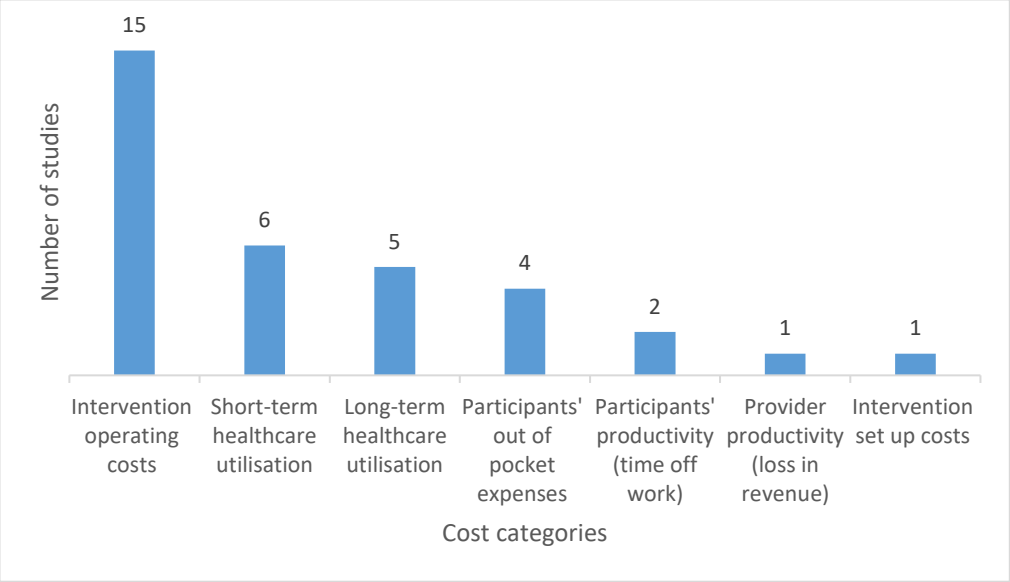


Figure 3. Cost categories identified across all 15 studies

Figure 3 shows that seven cost categories were identified across all 15 included studies. Like the findings in this review, Alayli-Goebbels et al. (2014) found the most common type of cost reported was the intervention costs, followed by healthcare costs. Participant out-of-pocket expenses and productivity losses appeared in only a small proportion of studies in this review and Alayli-Goebbels et al.'s (2014) review. Although most studies looked at both the direct and indirect costs of the interventions, only Edwards et al. (2013b) looked at the unintended productivity costs to the provider. More specifically, they examined whether the provider where the intervention was set (the leisure centre) experienced a loss in revenue, as a result of providing the intervention.

2.3.6. Challenge 4: Incorporating equity

The two former reviews found that authors did not routinely consider equity in their analysis (Alayli-Goebbels et al., 2014, Weatherly et al., 2009). Table 3 shows that all but one study (Shaw et al., 2011) included in the present review did consider equity. All but one study (Edwards et al., 2013b) did this implicitly, conducting subgroup analyses of the cost-effectiveness result (n=6) or targeting the intervention at a population deemed in need of intervention (n=8). Edwards et al. (2013b) were the only authors to explicitly discuss equity and to consider socio-economic status in their equity analysis. They did this by asking participants from areas of different levels of deprivation about how much they would be willing to pay to participate in the intervention of interest; thus informing the reader about participants' economic preferences. Notably this was an exploratory analysis and so the results were not incorporated in the CUA.

Table 3. Types of equity considered

Subgroup analyses of cost-effectiveness result	Campbell et al. 2015	Pre-existing condition
	Pavey et al. 2011	Pre-existing condition
	Anokye et al. 2011	Pre-existing condition
	Edwards et al. 2013; Murphy et al. 2012	Medical diagnosis
		Referral reason
		Adherence to scheme
		Gender
		Inequalities
		Age group
	Over et al. 2012	Age group
	Anokye et al.2012 ; Anokye et al. 2014	Age group
Intervention targeted at equity group	de Vries et al.2016	Frail older adults with mobility problems
	Leung et al.2012	Older adults
	Iliffe et al. 2014	Older adults
	Boehler et al. 2011	Older adults
	Maddison et al. 2015	People with ischaemic heart disease
	Elley et al. 2011	Females

	Young et al. 2012	Females
	Larsen et al. 2015	Latinas
Willing to pay question	Edwards et al. 2013; Murphy et al. 2012	Socio-economic status (level of deprivation)

Table 3 details the eight studies which targeted their intervention at a specific population group as well as the six studies that performed subgroup analyses of their cost-effectiveness result. Older adults was the most common equity subgroup targeted for intervention (de Vries et al., 2016, Iliffe et al., 2014b, Leung et al., 2012, Boehler et al., 2011), followed by females (Young et al., 2012, Elley et al., 2011). The most common subgroup analyses were on pre-existing condition/ medical diagnosis (Pavey et al., 2011b, Anokye et al., 2011, Edwards et al., 2013b, Campbell et al., 2015a) and age group (Edwards et al., 2013b, Over et al., 2012, Anokye et al., 2012). Edwards et al.(2013b) carried out seven types of equity analyses, all other authors conducted just one type. Furthermore, no studies attempted alternative equity analyses, such as an equity constraint or equity weighing analysis (Cookson et al., 2017).

2.3.7. New studies

The results of the rapid systematic scoping search are presented in Appendix A.2. In brief, four additional studies were identified as meeting the inclusion criteria of this review. Notably, one study (Gao et al., 2018) was an intervention targeting SB as an independent risk factor from PA. Furthermore, two studies (Gao et al., 2018; Harris et al., 2018) were both trial-and model-based economic evaluations, as the analysts had extrapolated their within-trial results a lifetime horizon.

2.4. Discussion

This review identified 15 economic evaluations of interventions that targeted physically inactive adults, and no economic evaluations of interventions that targeted sedentary adults (where SB was addressed an independent risk factor from PA). Like Abu-Omar et al's (2017) review of reviews which focuses on the results of economic evaluations, this present review identified economic evaluations on a limited range of PA interventions (Abu-Omar et al., 2017). Studies came from just four high-income countries, with over half (n=8) coming from the UK. This points to an important evidence gap in countries where economic evaluations are deemed appropriate. Examining a country's traditional beliefs around personal responsibility, efficiency and equity can explain why countries such as France and Germany are low users of economic evaluations and can in part explain why no studies in this review originated from these countries (Torbica et al., 2018). Regardless of cultural and institutional differences, globally health economists agree economic evaluations of preventative interventions are expected to have an important impact on future healthcare decision-making (ISPOR, 2018). In order to answer upcoming complex public health challenges, researchers need to go beyond clinical effectiveness methods and use a multidisciplinary suite of methods (Rutter et al., 2017) which includes economic evaluation. A prerequisite for this is an understanding on how key methodological challenges can be addressed.

2.4.1. Challenge 1: Attribution of effects

2.4.1.1. Modelling exercises

All ten trial-based economic evaluations in this review had a short time horizon; meaning they did not attempt to extrapolate or model the long-term impact of the intervention which could be used to informer longer term investment decision making. Any future reduction in incidence of NCD and premature mortality, attributable to PA and SB interventions, is unlikely to manifest until decades after the intervention has taken place. Yet, evaluating these interventions over the wrong timeframe means these interventions may appear ineffective or markedly less effective; they are at risk of not being appropriately prioritised by policymakers (Rutter et al., 2017). Curative interventions that rescue people from very poor health to better health will continue to be favoured, even if they are less cost-effective overall. Alayli-Goebbels et al. (2014) had previously suggested modelling as a way to extend the time horizon of trial-based studies, yet none of the ten trial-based studies in this review performed any modelling exercises. The challenges which can preclude extrapolation include the availability of data, and time and skills of the analyst (Squires et al., 2016).

2.4.1.2. Cohort studies

Campbell et al. (2015a) and Anokye et al. (2012) were the only two studies in this review to identify additional evidence to link up their short- and long-term effect estimate. The three other model-based studies claimed there was insufficient evidence to verify the accuracy of their assumptions (Pavey et al., 2011b, Anokye et al., 2011, Over et al., 2012). Notably, the cohort studies which Campbell et al.(2015a) and Anokye et al.(2012) draw on were published several years prior to the publication of the three other model-based studies. This suggests that the methodological challenge of ‘attribution of effect’ may be more dependent upon the analysts’ time and skills as opposed to the availability of data.

2.4.2. Challenge 2: Measuring and valuing outcomes

2.4.2.1. Cost-effectiveness and cost-utility analyses

This review found large inconsistencies in the types of outcomes measured and valued. There is no agreed classification system for PA outcomes (Abu-Omar et al., 2017) since the analysis of raw objective accelerometer data measuring objective PA levels is still in its infancy. Presenting a limited range of results can reduce the applicability of the study’s findings to other policymakers. Authors’ views also differed firstly on whether short-term QALYs should be included in the economic analysis, secondly on whether a short-term QALY gain represented a one-off gain in mental health or general functional health. Presently, within the economic literature the responsiveness of the EQ-5D-3L to detect important differences in the severity of health is being challenged, and had led to the development of the EQ-5D-5L, which measures health on five levels as opposed to just three (Glick et al., 2014). This review has shown that outcomes used in PA studies are diverse; therefore, there is a need for analysts to agree on a consistent outcome that best captures the objectives of a PA intervention.

2.4.2.2. Cost-benefit and cost-consequence analyses

No studies in this review performed a CBA and just two presented a CCA alongside their full economic evaluation. There is a lack of CBAs in other public health areas. Hill et al. (2017) and

Alayli-Goebbels et al. (2014) identified a small proportion of studies (n=1 and n=8 respectively) who reported conducting a CBA, but due to insufficient reporting gained limited insight into how these were performed such as how outcomes had been monetised (Hill et al., 2017, Alayli-Goebbels et al., 2014). Likewise, four studies claimed to be CBAs in the review by Weatherly et al.(2009), but after further assessment were re-classified as CCAs (n=3) and a CEA (n=1). Although classified as a partial-economic evaluation, CCA is a useful alternative to CBA since all relevant costs and consequences can be presented to the reader in the form of an inventory, rather than simplified into a single outcome measure or index as is the case in CEA and CUA, respectively. If an outcome is deemed relevant to the reader, they can reanalyse the data quantified in the CCA. However, CCA puts more onus on decision makers than CBA or CUA, as it does not roll outcomes into a summary measure that can be compared to a decision rule. An example of a decision rule in the UK is: invest where the incremental cost-effectiveness ratio is less than £30,000 per QALY (NICE, 2014a).

2.4.3. Challenge 3: Identifying intersectoral costs and consequences

2.4.3.1. Inconsistent perspectives

The three most common perspectives stated were the health system, payer and societal perspectives. These match the three most commonly reported perspectives in the broader field of economic evaluation (Husereau et al., 2013). Only Edwards et al. (2013b) conducted their analysis from the public sector perspective, a perspective recently recommended in the UK reference case (NICE, 2014a). That said, Edwards et al. (2013b) did not incorporate participant costs in their CUA, only through an exploratory analysis. Only three studies considered the cost to the participant, which is not surprising since the health sector perspective was the most commonly stated perspective. Participant and voluntary sector costs are deemed important, but previously have not been routinely captured (Weatherly et al., 2009).

It was found that even economic evaluations stated the same perspective did not always include the same costs and consequences. This is likely to be because there is a lack of standard definitions for the various perspective types (Husereau et al., 2013). Even where there are examples of standard definitions, such as those proposed by the Second US Panel on Cost-Effectiveness in Health and Medicine (Sanders et al., 2016), not all economists agree with their definitions, and furthermore the definitions may not be applicable to other countries since there are distinct features of each health system (Torbica et al., 2018). For instance, deciding what costs and consequences to capture within a societal perspective is a normative question, requiring the analyst to make social value judgements (Drummond et al., 2015b). This is an important issue, since the exclusion of relevant consequences can lead to an underestimation of cost-effectiveness whilst the exclusion of relevant costs can lead to an overestimation of cost-effectiveness (Hill et al., 2017).

2.4.3.2. Cost categories identified

The cost categories identified in this review match the five cost categories (healthcare services, intervention costs, patient and family costs, lost productivity costs, future costs) identified as most relevant for inclusion in economic evaluations, by health economists who recently took part in a

cross-Europe Delphi study (van Lier et al., 2017). This suggests analysts' choice in costs in this review align with analysts in the more general field of economic evaluation. It should be noted however that there was a difference in one of the categories, as family costs were not identified as a relevant cost category in the studies from this present review. Just two trial-based studies included absenteeism in their study; similarly only two of the model-based studies included it in their CCA. It continues to be debated in the literature as to whether absenteeism is an outcome of cost-offset, and thus whether it should be included in the numerator or denominator part of the incremental cost-effectiveness fraction (Drummond et al., 2015b).

2.4.4. Challenge 4: Incorporating equity considerations

2.4.4.1. Presenting results by subgroups

Equity impact analysis can be as straightforward as presenting cost-effectiveness results by equity subgroups (Hill et al., 2017, Alayli-Goebbels et al., 2014, Weatherly et al., 2009). Six studies in this review presented an equity impact analysis (Campbell et al., 2015a, Pavey et al., 2011b, Anokye et al., 2011, Anokye et al., 2012, Over et al., 2012, Edwards et al., 2013b). The most common subgroup analysed was individuals with pre-existing medical conditions, nevertheless this analysis was performed in just four studies (Campbell et al., 2015a, Pavey et al., 2011b, Anokye et al., 2011, Edwards et al., 2013b). Furthermore, only one study (Edwards et al., 2013b) conducted more than one type of equity subgroup analysis. These findings suggest analysts are not performing equity analyses in a comprehensive nor consistent manner. Weatherly et al. (2009) outlined socio-economic status as an important under-researched equity issue in economic evaluations, however only one study in this review researched socio-economic status by asking participants about their willingness to pay for an intervention component (Edwards et al., 2013b). Incorporating equity into decisions on PA and SB interventions is especially important, since it is amongst the lower socioeconomic groups where physical inactivity is greatest (Organisation for Economic Cooperation and Development, 2015).

2.4.5. New studies

Overall, the four studies published since March 2017 did not change the narrative of this review since there remains a dearth of economic evaluations in the field of PA and SB. What the studies have demonstrated is that firstly, there is an indication that health economic methods have begun to be applied to targeted SB interventions (Gao et al., 2018). Secondly, that it is feasible and informative to extrapolate beyond the trial (Gao et al., 2018, Harris et al., 2018).

2.4.6. Strengths and limitations

This is the first systematic review conducted since 2009 to review the methods used in economic evaluations of interventions targeted at physically inactive individuals, and the first systematic review to search for economic evaluations targeting SB as an independent risk factor from PA. This review included comprehensive literature searching and a rigorous methodology in line with the PRISMA guidelines (Moher et al., 2009). Economic evaluations aim to inform resource allocation decisions (Drummond et al., 2015b). Previous reviews have demonstrated that key methodological challenges preclude economic evaluations in the field of public health from achieving this aim (Weatherly et al., 2009, Alayli-Goebbels et al., 2014). By focusing on PA and SB, this review has

been able to not just provide an overview on whether or not the four key methodological challenges have been addressed in the last decade, but crucially explain in greater depth the methods performed in those few studies where progress has been made.

More specifically, progress has been observed in the 14 studies which have considered equity in their analysis (Table 3) and the small proportion of studies where: the long-term model presented has been informed by robust epidemiological evidence (Anokye et al., 2012, Campbell et al., 2015a); all important and relevant costs and consequences have been outlined to the reader in the form of a CCA (Pavey et al., 2011b, Anokye et al., 2012); and/or a multi-sector perspective has been selected (Edwards et al., 2013b). An output from the narrative synthesis of this review is a number of recommendations (as outlined in Table 2.4) explaining how analysts can continue to make progress towards addressing the four methodological challenges. Although, the comprehensive search strategy only goes up to March 2017, a rapid systematic scoping search is presented which highlights four new empirical studies. Two of these studies (Harris et al., 2018, Gao et al., 2018) support the recommendations emerging from this review in terms of linking up the intermediate evidence with longer term policy relevant outcomes.

It was not within the scope of this research to review the methods used in population-level interventions such as national policies or media campaigns. It would therefore be useful for future reviews to explore how economic evaluations are being carried out within this area. In addition, this review focuses on the methods conducted in full economic evaluations and so there is scope to review the methods used in partial evaluations. Nevertheless, full economic evaluations are deemed more informative than partial evaluations, and so it would have been expected that analysts would conduct for instance, a CCA alongside their full economic evaluation, as was done in two studies (Pavey et al., 2011b, Anokye et al., 2014) in this review.

2.4.7. Recommendations

Table 4 presents a list of recommendations for researchers and users of economic evaluations from a variety of disciplines (health economics, public health, PA etc) to refer to when designing, analysing and appraising economic evaluations of targeted PA and SB interventions.

Table 4. Key recommendations for future economic evaluations

Challenge	Recommendation	Explanation
Challenge 1. Attribution of Effects	*Modelling*	It is necessary for public health researchers to invest time in reviewing the existing evidence base and develop novel modelling skills. Best practice guidelines state well established published models are preferred to those developed specifically for a trial (Ramsey et al., 2015). If skill and time permits, analysts can draw on the structure of the published models (Campbell et al., 2015a, Anokye et al., 2012) identified in this review and adapt them according to the local decision-making context.
Challenge 2. Measuring and valuing outcomes	*Cost-consequence analysis*	There is a need for further methodological developments in the monetisation of effects in CBAs (Drummond et al., 2015b, Sanders et al., 2016). In the meantime, it is deemed more appropriate to conduct a good quality CUA which may be of a narrower perspective, than a poor quality CBA which captures a broader perspective (Hill et al., 2017, Weatherly et al., 2009). In order to report on multiple outcomes which extend beyond health, a CCA or impact inventory conducted alongside a full economic evaluation is recommended (NICE, 2014a, Sanders et al., 2016). If the word limit in journals precludes authors from presenting a CCA in the main manuscript, they should present this information in the online supplementary material.
Challenge 3. Identifying intersectoral costs and consequences	*Multi-agency public sector perspective + participants perspective*	Three studies in this review omitted costs, which would typically be deemed relevant to their stated perspective, and two studies did not report their perspective. It is imperative for analysts to describe and justify the costs and consequences, which they have deemed relevant for their chosen perspectives (Husereau et al., 2013). Inevitably different assumptions on what costs and consequences are included in the analysis leads to different results (Sanders et al., 2016). Furthermore, future studies should aim to present at least two types of perspectives and conduct a CCA or impact inventory alongside their CUA or CEA in order to present the various relevant costs and consequences to the various relevant sectors (Weatherly et al., 2009, Sanders et al., 2016,

		Alayli-Goebbels et al., 2014). A multi-agency public sector perspective where costs and consequences are presented in their disaggregated form (i.e. in a CCA) for each sector is preferred over stating a societal perspective (Drummond et al., 2015b, Hill et al., 2017). It is also recommended that future studies, specifically trial-based studies, capture economic information on time, travel and out-of-pocket expenses incurred by the participant. The participant's perspective is important as the amount of time and expenses they invest may have an impact on the participant's uptake, adherence and overall acceptability of the intervention.
Challenge 4. Incorporating equity	*Equity impact analysis*	Analysts should present costs and consequences explicitly in their disaggregated form for various equity groups, so policymakers can start to build a better picture on which population groups gain and lose from a specific decision (Hill et al., 2017). From here, analysts can conduct an equity impact analysis. This type of analysis is deemed easier than conducting equity constraint or equity weighting analysis (Hill et al., 2017). The equity effectiveness loop framework (Welch et al., 2008) and PROGRESS-Plus framework (O'Neill et al., 2014) are recommended to help analysts consider, in a structured way, which equity factors may be relevant to their study (Alayli-Goebbels et al., 2014, Welch et al., 2017).

2.5. Conclusions

A focus on the key methodological challenges in economic evaluations is important, as they can impact on the derived cost-effectiveness result, which ultimately can impact on a policymaker's resource allocation decision. As economic evaluation is a rapidly developing field (Drummond et al., 2015b) this systematic review has provided an important update on the most recent methods used in targeted PA interventions. The review has also highlighted there is a scarcity of economic evaluations for targeted SB interventions. Importantly, this review makes it explicit to policymakers and researchers from the varied disciplines in which PA and SB falls under, that there are still key methodological challenges that need further attention. This review has highlighted that methodological choices vary widely not just between countries but also within them. Ultimately, these analyst-based choices affect the results presented and subsequent resource allocation decisions made. A recent consensus statement has called for collaboration across the disciplines to develop guidance specific to the context of economic evaluations of PA interventions (Davis et al., 2014). To date, no guidelines have been developed to address this need. The examples of methodological development identified from the studies in this review and the resulting review recommendations can be used to inform future guidelines and their supplementary materials. In particular, they will be used to develop an initial outline of the framework, which will be presented in Chapter 3.

Chapter 3: Development of an initial systematic framework for economic evaluations of individual-level PA and SB interventions

3.1. Chapter aim

The systematic review presented in Chapter 2 demonstrated how economic evaluations are being performed inconsistently across interventions of the same nature including those from the same country. Accordingly, a framework is needed to support multidisciplinary researchers conducting trial-and model-based economic evaluations (Davis et al., 2014). The aim of this chapter was to develop an initial multidisciplinary framework which can be used by researchers including those who have limited or no specialist training in health economics and would like to conduct economic evaluations of individual-level PA and SB interventions. The framework may also be of use as a guiding framework for health economists seeking a standardised approach to economic evaluation of individual-level PA and SB trials. More specifically, the framework will be a guidance tool highlighting ideal practice to the user based on the literature and standard practice from the fields of: health economics, public health, behavioural science, PA, SB and trial methodology. In order to assess the practicality of the guidance framework, it has been piloted in two trials: (1) a pragmatic quasi experimental trial of a co-developed PA on Referral Scheme (Co-PARS) (Chapter 4); (2) a randomised controlled trial aiming to help mainly desk-based workers to Sit Less and Move More at work (SLaMM trial) (Chapter 5).

3.2. Framework development

3.2.1. Procedure

Four key steps were taken to develop the framework, these included:

1) Structuring the framework by drawing on the 10 generic methodological steps from the economic evaluation quality assessment checklist applied in Chapter 2.
2) Drawing on the results from Chapter 2 to recommend how the four key methodological challenges could be addressed.
3) Identifying data collection tools used in existing economic evaluations.
4) Multidisciplinary team meetings

3.2.2. Structuring the framework

The structure of the framework draw on Drummond's quality assessment framework (Drummond et al., 2015b) which had been applied in Chapter 2. This framework was selected as it names the key methodological steps for the conduct of generic economic evaluations. The framework was adapted so as costs and outcomes were itemised separately. This was because Chapter 2 had illustrated that the recommended approaches for identifying, measuring and valuing costs and outcomes would differ in a number of ways. Consequently, this led to three additional items being included in the framework. A further three items were also added based on key recommendations identified in the wider methodological literature. These additional methodological items considered, the need to: (1) adjust for baseline imbalances in costs and health-related quality of life values (Manca et al., 2005, Franklin et al., 2019); (2) consider equity (Weatherly et al., 2009); and (3) adhere to the consolidated health economic evaluation reporting guidelines (Husereau et al., 2013). Overall, the structure of the framework comprised of 16 items.

3.2.3. Addressing the four key methodological challenges

Chapter 2's findings offered insight into how four key methodological challenges could be addressed in the context of individual-level PA and SB trials. Methodological approaches that were frequently reported and/or were reported in good quality studies, were deemed most important. In summary, based on Chapter 2's findings the following types of recommendations were made: time horizon for the analyses, cost categories to assess, valuation techniques for measuring and valuing the outcomes, and lastly an approach for incorporating equity into the assessment.

3.2.4. Identifying data collection tools

To produce a framework that can be applied by researchers wishing to explore cost-effectiveness, it was necessary to provide specific guidance on what cost items would be captured for each cost category and what measurement tools would be used to capture these items. The initial intention was to revisit studies from the review in Chapter 2 and perform a more granular analysis of the data collection methods used in studies that met items 1, 4, 5 and 6 from the quality assessment checklist (Appendix A.4). These four quality assessment items were deemed most relevant to the purpose of the synthesis, as they consider the perspective stated, and how costs and effects were identified, measured and valued (Drummond et al., 2015b). Five trials and one modelling study met the four specified inclusion items (these studies are listed in Appendix B.1). It should be noted that two modelling studies (Campbell et al., 2015b, Anokye et al., 2011) also met the inclusion items for nine of the items, but not item 5, as it was not possible to know whether appropriate physical units had been measured accurately. Nonetheless, the two studies referenced the same microcosting study (Isaacs et al., 2007) and so the cost items used in that study were reviewed (Appendix B.1). By large, it was found that the data collection methods in the studies from Chapter 2, were poorly reported and so it was not possible to draw useful information from these studies. Accordingly, the database of instruments for resource use measurement (DIRUM) (Ridyard and Hughes, 2012) was drawn on to identify appropriate measurement tools to recommend in the framework. Just one tool was identified from the database to measure participant costs which was a generic tool for any disease area. No tools were identified for measuring productivity or intervention costs for any of the disease areas listed in the database.

3.2.5. Multidisciplinary team meetings

Several iterations of the framework were drafted and reviewed through my supervisory team that comprised of senior researchers in PA and SB (n=2), public health (n=1) and public health economics (n=1). I had monthly meetings with my supervisors at the development phase of my framework. At these meetings, I shared: (1) key observations I had made regarding the methods used in the studies included in my systematic review; and (2) key reflections I had made regarding the relevance of additional methodological papers I had come across in the health economic and public health literature. The aim of these meetings was to informally discuss and reflect on the relevance, importance and practicality of the approach I was developing for the framework.

During the development stage of the framework, I was successful in gaining an international mobility award which enabled me to draw on additional health economic expertise at Deakin

University. I undertook a three-week study placement that included two one-hour face-to-face meetings with three senior health economists who had recently been involved in conducting an economic evaluation of an individual-level SB trial. I asked these experts to share their experience on conducting an economic evaluation in the field of public health. For instance, I asked them how they typically identified intervention activities and how they incorporated productivity into their studies. In addition, I presented the findings and recommendations from my systematic review to the whole health economic group at Deakin University. I invited the group to ask me questions in order to yield insight into what they thought about my recommendations. The main discussion was around equity. There was consensus amongst the group that equity was important to include in public health economic evaluations, but most felt it was not yet standard practice. I reflect further on these key discussions in the reflection boxes in the section 3.3.

3.3. Initial outline of a framework

An initial outline of a framework was developed, this is presented in Table 1 in the form of questions Section 3.3 provides recommendations on how the framework can be applied in practice. Alongside the recommended items are reflections which explain why specific items were included in order to ensure the design of the framework is transparent. sections. Box 1 below introduces the main types of economic evaluation.

Box 1. Definitions: Economic Evaluation Techniques

Full economic evaluation: There are three established techniques for "full" economic evaluations: cost-effectiveness analysis (CEA), cost-utility analysis (CUA) which is often called CEA or referred to as a special type of CEA, and cost-benefit analysis (CBA). The economic evaluation handbook by Drummond et al. (2015) includes further detail around these techniques, including their theoretical underpinning. In summary, the important difference to note is the way in which the techniques include and value outcomes in the analysis. The main difference between conducting a CEA or CBA in practice relates to how many outcomes you include in the analysis and how you choose to value your outcomes. CEAs incorporate just one outcome measure (as well as costs) which is reported in non-monetary units. By contrast, CBA can include several outcomes, which are converted into monetary values. *Partial-economic evaluation***:** Cost-consequence analysis (CCA) is a partial-economic evaluation as costs are not compared to effects. Instead, the technique involves listing all costs and effects in their natural units but does not attempt to aggregate the costs and effects into a summary statistic.

Table 1. Summary of items to consider when conducting an economic evaluation of an individual-level PA and SB intervention

Items*	
1	What components make up a well-defined study question?
2	What does a comprehensive description of the comparators look like?
3	What does an appropriate study design look like?
4	What costs are important and relevant?
5	What effects are important and relevant?
6	How and when can costs be measured?
7	How and when can effects be measured?
8	How can costs be valued?
9	How can effects be valued?
10	How can costs and effects be discounted to a present day value?
11	What summary statistics can be presented?
12	What adjusted analyses can be performed?
13	What equity subgroups can be considered?
14	What uncertainty analyses can be performed?
15	How can the results be interpreted?
16	How can trial-based economic evaluations be reported?

** Adapted from the methodological checklist outlined in the Drummond et al. (2015) quality assessment appraisal checklist for economic evaluations*

3.3.1. Item 1. What components make up a well-defined study question?

The five pieces of information listed 1.1-1.5 below can help with defining your study question. The reflection boxes help explain why these five pieces of information are advised.

3.3.1.1. Item 1.1. Comparison of both costs and effects for at least two groups

It is advisable that you confirm that both costs and effects of at least two comparator groups are being evaluated. Often this might be an intervention vs. a ‘no intervention’ or ‘treatment as usual’ comparator.

Reflection on item 1.1

This framework starts by recommending the analyst confirms that their study has two fundamental characteristics: (1) a comparison of at least two comparator groups (where the term comparator group is synonymous with the term intervention, service and strategy); and (2) a comparison of both costs and effects. It was through the process of developing the eligibility criteria for my systematic review that I discovered that the health economic literature define an economic evaluation as having these characteristics and as being one of the following techniques: CEA, CUA, CBA and CCA (Drummond et al., 2015b). Initially I was surprised to learn that SROI, an economic technique I had used as a Public Health Assistant, was not reported as a main evaluative method. However, the definition helped me recognise that the SROI evaluations I had performed in my previous job were limited to one comparator group (they were typically 'before and after' evaluations) and therefore did not meet the study design criteria to be regarded as an economic evaluation. For this reason, I have not recommended SROI for my initial framework. In terms of clinical effectiveness trials, although they typically have two or more comparator groups but they typically only compare effects.

3.3.1.2. Item 1.2. CEA as the primary analysis

It is helpful to specify the primary analysis by reporting: (1) the type of economic evaluation you will conduct (see Box 1); and (2) the country and decision-maker (perspective) your results are intended for. Before specifying your primary analysis check whether your country has specific guidelines on what the preferred primary analysis is. If your country does not specify which analysis technique is preferred, it is advised that a CEA is conducted as opposed to a CBA. Secondly, it is advised that you conduct the analysis from a broad perspective including all stakeholders who have the potential to be affected by the intervention rather than just the perspective of healthcare organisation. One way to help you identify which additional organisations and individuals may be important, is to identify who is involved in paying, providing and/or setting up the intervention of interest.

Reflection on item 1.2

The type of economic evaluation recommended for the primary analysis is a CEA. All studies identified in my systematic review (Chapter 2) had conducted a CEA (or CUA) with no studies performing a CBA. In the discussion section of my systematic review, I indicate that this may be due to the practical challenge of conducting a CBA (e.g. the difficulty of assigning monetary values to non-health outcomes). I did not have any experience or practical examples from the literature on how to deliver a CBA in practice, therefore I do not recommend it in this initial framework.

In my review, I also suggest that the choice in methods is likely to be also due to the fact many countries have guidelines or a 'reference case' which states the country's preferred analysis technique. Through my health economic reading, I learnt that the aim of a reference case is to improve comparability of studies reported from the same country (ISPOR, 2019). Prior to conducting the systematic review I had not come across the term 'reference case' as this concept is not referred to within the clinical effectiveness evaluations. An example from the UK's 'reference case' is that it recommends a specific type of CEA, a CUA, is performed for the primary analysis (NICE, 2014).

3.3.1.3. Item 1.3. CCA as a secondary analysis

CEA produce aggregate summary outcome statistics, therefore it is recommended that a cost consequence analysis (CCA), also known as an impact inventory, is conducted alongside the primary analysis (Sanders et al., 2016). This is so all relevant costs and outcomes assessed can be interpreted separately. As interventions can be delivered in public or non-public sector settings, the CCA may include costs and effects deemed important to both the public and private sector agencies who are involved in paying, providing and/or setting up in the intervention. In addition, it is recommended that the participant's perspective (also known as the private individual's perspective) is captured in the CCA.

Reflection on item 1.3

CCA is considered a partial-economic evaluation since costs and effects are not compared in order to produce a summary cost-effectiveness statistic (Drummond et al., 2015a). In my systematic review (Chapter 2), I identified two studies that reported a CCA alongside their primary analysis (CEA or CUA). I saw how the disaggregated format of the costs and effects presented in these CCA made it clear to identify which costs and effects were relevant to which stakeholder. In addition, my review identified four economic evaluations, which had included participant costs in their analysis. Even though many country's guidelines including the UK do not recommend the inclusion of participant costs (ISPOR, 2019, NICE, 2014a) I felt the participant's perspective was important to include in the CCA. This is because the behavioural science literature argues that the participant's acceptability of (perspective on) an intervention can provide influence the success of an intervention (Michie, Atkins, & West, 2014).

3.3.1.4. Item 1.4. Time horizon

It is recommended that the primary analysis is conducted over the trial follow up period (trial time horizon). If there is sufficient data, time and expertise, a decision model could also be conducted. The model could assess the costs and effects over the rest of the participant's life (lifetime horizon).

Reflection on item 1.4

When resource use is collected from a single trial, best practice guidelines state that a trial-based economic evaluation is always conducted before any modelling and therefore the first analysis should be based on a short time horizon which will be the trial follow up period (Glick et al., 2014). My systematic review (Chapter 2) indicated that trial-based economic evaluations that do not extrapolate beyond the trial follow up period are more commonly performed than economic evaluations that do extrapolate. Nonetheless the epidemiological literature indicates that the greatest potential benefits of increasing your weekly PA levels are unlikely to accrue until decades after the intervention has taken place. My systematic review identified a small number of studies which estimated the long-term cost-effectiveness of the PA interventions by linking up short-term surrogate outcomes with published epidemiological evidence. This made me aware that there are epidemiological data available within the literature in order to build evidence-based decision model for PA interventions.

3.3.1.5. Item 1.5. Target population and subgroups

It is helpful to confirm the target population for the economic evaluation. The target population for the trial-based economic evaluation could be the participants recruited for the clinical effectiveness trial. The modelling analysis could include a broader population if there is sufficient evidence from other studies published. In addition, confirm which equity subgroups will be considered, at a minimum consider: age, sex, socioeconomic status and medical condition.

Reflection on item 1.5

Socioeconomic status, sex, age and pre-existing medical condition were the most common types of equity subgroups identified in my systematic review (Chapter 2). The public health research I was involved with before I undertook this PhD had made me believe that socioeconomic status is likely to be an important equity subgroup for public health trials.

3.3.2. Item 2. What does a comprehensive description of the comparator groups look like?

Key information required to describe the intervention groups is likely to be provided in the trial's protocol and the CONSORT flow diagram if available. If you feel the protocol and CONSORT flow diagram does not provide sufficient detail, it is helpful to heck your interpretation of the comparator groups with those leading the clinical effectiveness trial. After the intervention has been delivered, it is also advised that you check with the staff who delivered the intervention whether any additional intervention operating and/ or set up costs were incurred. These could be costs that have not been accounted for in the trial protocol or CONSORT diagram. In addition to describing the intervention of interest, it is recommended that you describe what usual care is, as usual care may vary geographically and/or by organisation. The description of the comparator groups can be used as

the first step to populating a microcosting tool (Table 5). The tool can be populated by the research team and/or staff delivering the intervention.

Reflection on item 2

I was interested in how the studies I included in my review had calculated the intervention costs, since the descriptions of the interventions in the articles were brief and the authors did not comment on the methods used to identify the resource quantities. During my study placement with Deakin University's Health Economics group, I had the opportunity to explore the costing of complex public health interventions with three senior academic health economists. They discussed that they typically use the CONSORT diagram and study protocol to identify the multiple intervention activities and the resources required to deliver them. The idea to check whether any additional unexpected intervention costs had been incurred after the delivery of the intervention came from one of the studies in my review (Edwards et al., 2013b). This study conducted telephone interviews with the intervention providers after the intervention had been delivered.

3.3.3. Item 3. What does an appropriate study design for trial-based economic evaluations look like?

In order for the results of your economic evaluation to be applicable to 'real world' decision maker who need to make inevitable decisions around resource allocation, it is recommended that the economic data you use in your analysis is derived from a pragmatic trial. If possible, a randomised controlled trial (RCT) is recommended for the trial design. Furthermore, it is recommended that the Intention-to-treat (ITT) principle is followed for the primary analysis of the trial data, whereby participants are analysed according to the same group they were assigned to, even if they do not adhere to their allocated intervention. If there is a more than 10% of data missing, then a complete case analysis is recommended.

Reflection on item 3

Pragmatic trial

It was through my reading of one of the health economic handbooks that I became aware that there was consensus amongst the health economic community that economic evaluations should be delivered alongside pragmatic trials (Drummond et al., 2015a). I think my anthropology background also helped me recognise why pragmatic trials were required for economic evaluations of behaviour change trials (e.g. changing PA and SB levels). Anthropology is the study of complexity. I therefore had experiencing of studying complexity which is important for complex interventions. There is increasing recognition in the trial literature that there is a need to document and understand complexity of interventions, in order to understand why some interventions fail to be implemented into the 'real world' (Moore et al., 2015).

Intention-to-treat principle

A key feature of pragmatic trials is that it necessitates that study participants are analysed in accordance to the comparator group they are assigned to at baseline, even if the participant does not adhere to the protocol or changes groups during the trial. This trial design feature is known as the intention to treat (ITT) principle. I verified that the ITT principle applied to trial-based economic evaluations by consulting good practice guidelines provided by the International of Society for Pharmoeconomics and Outcomes Research (ISPOR) (Ramsey et al., 2015).

Complete case analysis

Complete case analysis is reportedly the most common way to analyse incomplete datasets in trial-based economic evaluations (Noble et al., 2012). I tried to understand why multiple imputation was not used more commonly for economic evaluations. I came across a methodological study which suggested that this may be due to the lack of guidance for addressing missing data in the context of economic evaluation (Leurent et al., 2018). I used the literature to inform my recommendation on what the cut-off point would be for missing data. I identified a study which stated that for multi-item measurement tools where only a small proportion of the data is missing (less than 10%) it is deemed acceptable to impute the mean of each group for participants missing an item (Eekhout et al., 2014).

3.3.4. Item 4. What costs are important and relevant?

Important and relevant costs categories are likely to be influenced by the country and/or audience of your economic evaluation (as discussed in item 1). The description of the comparator groups as described in item 2 will help you judge what costs are relevant. At a minimum, if conducting your analysis from a multi-agency public sector perspective (as recommended in item 1) it is helpful to consider including the following perspectives and associated cost categories:

1. *Payer's perspective*: intervention costs, which could include the setting up (organising) and operating (delivery) costs.
2. *Provider's perspective:* any additional intervention operating and setting up costs, not accounted for before the trial is conducted
3. *Health and social care perspective:*
    - Short-term primary healthcare activity including: consultations with the GP, practice nurse and allied health professionals, and medications prescribed in primary care
    - Short-term secondary healthcare activity including: emergency, outpatient and inpatient visits
    - Long-term healthcare activity: secondary data can be used to estimate potential future treatment costs in a decision model. For PA it is recommended that future treatment costs for the following diseases are considered: Type 2 diabetes (T2D),

stroke and Coronary Heart Disease (CHD). For SB interventions, at a minimum treatment costs for T2D.

4. *Participant's perspective:* out of pocket expenses such as clothing and travel costs, and time costs such as loss in leisure time to attend intervention activities. In addition, an exploratory analysis is recommended whereby participants are asked about their willingness-to-pay for a PA or SB intervention.
5. *Employer's perspective*: losses and gains in productivity. This will be particularly relevant if the intervention is set in the workplace and/or requires the participant to lose time from work in order to participate in the intervention.

Reflection on item 4

Identification of cost categories

A key finding from my systematic review (Chapter 2) was that studies conducted from the same perspective (e.g. societal perspective) included different cost categories in their analysis. This even applied to studies conducted in the same country. The health economic literature reports that, internationally, the process for identifying cost categories and items for each perspective type (e.g. healthcare, societal, payer) is in part analyst-dependent as international standardised definitions do not exist (Husereau et al., 2013). The studies in my review did not state why they had included or excluded specific cost categories. As a result, I found it difficult to recommend a systematic approach for identifying all important cost categories. The approach I therefore recommend is a comprehensive approach as it lists all the cost categories which could be included in a study. This list was identified from seven studies from my review which met items on my quality assessment checklist relating to the reporting of study perspective and costs (Table 1 in Appendix B.1 lists the seven studies and their cost categories used). The cost categories I identified through the seven studies included: intervention operating costs, intervention setup costs, immediate healthcare utilisation, future healthcare utilisation, participant costs and productivity costs included in this framework. At the time of conducting this assessment of cost categories, a Delphi study was published with similar findings to my own (van Lier et al., 2017). The Delphi study identified five key cost categories: intervention costs, healthcare costs, patient and family costs, lost productivity costs and future costs.

Long-term (future) healthcare costs

In addition, I conducted a content analysis of the discussion sections of the 10 studies included in my systematic review which had not included future (long-term) costs in their analysis. I wanted to explore whether the authors saw this as a limitation of their study. I found that eight of the 10 studies reported that the exclusion of long-term costs was a limitation of their analysis. The two studies which did extrapolate their single trial-based results in order to consider the future costs conducted modelling and drew on pre-existing models (Gao et al., 2018, Anokye et al., 2018). It also saw that one author recommended

that existing models are drawn on to develop new decision models (Edwards et al., 2013b).

3.3.5. Item 5. What effects are important and relevant?

As discussed in item 1.2, many countries have guidelines (or a 'reference case') which specifies what the preferred outcome measure is for the primary analysis. It is therefore recommended that you consult your country's guidance. Guidelines can be retrieved from the ISPOR webpage on pharmacoeconomic guidelines from around the world: https://tools.ispor.org/peguidelines/ (ISPOR, 2019). If no guidelines are available for your country of interest, it is recommended that it is recommended that a single generic measure of health is used as the primary outcome measure (e.g. the QALY). Additional wider effects of interest may have also been identified by the trial team working on the clinical effectiveness evaluation. All effects for which it is not possible to assign a monetary value, can be presented alongside the cost data in the secondary analysis, the CCA.

Reflection for item 5 (reflections in item 1.2 also relevant to this item)

The most common primary outcome measure included in my systematic review (Chapter 2) was the QALY. QALYs are the most commonly used outcome measure in the literature and have fewer measurement problems compared to other outcomes such as DALYs. Although some studies in my review reported a measure of PA as the primary outcome, the outcome was reported in different units (e.g. one minute of PA, one person achieving 150mins of PA per week).

3.3.6. Item 6. How and when can costs be measured?

3.3.6.1. Item 6.1. Intervention operating and setting up costs

Prospective data collection is preferred as it is expected to be more accurate since it does not rely on participant/ staff/ researcher recall. Intervention operating and setting up costs can therefore be recorded by the research team during the trial using the microcosting tool (Table 5). The tool provides examples of the types of costs typical of individual-level PA and SB interventions such as staff type and time, equipment and capital equipment. The tool can be applied in Excel or similar software. The comprehensive descriptions of the new and existing interventions reported in item 2 can be used to support the microcosting exercise. It is helpful to engage the trial protocol, CONSORT flow diagram and wider research team in this process.

Reflection on item 6.1.

Three studies in my systematic review (Chapter 2) reported generating their intervention resource use estimates through study records (Iliffe et al., 2014a, Elley et al., 2011, Isaacs et al., 2007). Nevertheless, the authors did not provide examples on the tools or templates used to document these records. Another study from my review reported using a budget breakdown to estimate costs, which they retrieved from the organisation who

was paying for the intervention (Edwards et al., 2013b). It was not clear what tool or template was used by the local authority to document the intervention costs.

I had come across the DIRUM database through my reading of one of the health economic handbooks (Drummond et al., 2015a). As discussed in my systematic review, DIRUM is a repository of papers about resource use and cost measurement (Ridyard and Hughes, 2012). I searched the DIRUM repository to see if I could find a tool or template to help me document intervention costs in a systematic and comprehensive way- I could not identify any. It is possible that this is because tools used by health economists to evaluate complex lifestyle interventions may not have been validated and shared yet as DIRUM is a relatively new initiative and microcosting methods are underdeveloped (Frick, 2009). Nevertheless, there is recognition that microcosting is becoming increasingly important for newly developed complex multi-component interventions (e.g. individual-level PA and SB interventions) since it is likely these interventions have not been assigned an aggregate cost that is available in the published literature (Glick et al., 2014). In the absence of a validated microcosting tool, I developed my own novel microcosting tool in order to document interventions costs for the Co-PARS and SLaMM trial (Table 5). The structure of my tool in terms of the variables which have been included, were informed by the NHS reference cost structure (NHS Improvement, 2018). The content of the tool for the types and descriptions of the cost items (e.g. staff type, printing, room hire) was based on my content analysis of the cost items reported in the studies in my systematic review. More specifically, I analysed the method and result sections of seven studies which had adequately reported what cost items they measured for each cost category (see Table 1 in Appendix B.1).

Table 5. Microcosting tool to record operating and setup costs

Payer's perspective						
Intervention operating/ setup costs	Name and description of cost item	Average quantity of cost item	Average quantity of time, if applicable	Total Quantity	Unit Cost	Total Cost
Staff's time	Staff type (including details on: qualification level and/or grade) and time for the following activities: training, travel, preparation, delivery and clear-up					
Equipment	Printing					
	Physical materials					
	Promotional materials/ advertisements					

	Refreshments					
	Study specific software					
	Staff clothing					
	Home working facilities					
	Postage					
	Stationary					
	Phone costs					
Fixed	Private room					
	Non-study specific IT equipment					
	Non-study specific software					
	Staff overheads					
	Room hire					

3.3.6.2. Item 6.2. Additional intervention operating and setup costs

These costs can be captured at the end of the trial by interviewing staff from the settings where the interventions are being delivered. The interview schedule provided in Appendix B.1. is recommended.

Reflection on item 6.2.

The recommended schedule for interviewing relevant staff/ stakeholders is presented in Appendix B.1. I identified just one study in my systematic review (Chapter 2) which captured additional operating and set up costs, and they had done this using telephone interviews with staff (Edwards et al., 2013b). Nonetheless, the study did not include an example of the interview schedule they used to capture this cost type. I therefore searched the DIRUM database to see if I could identify a template for an interview schedule that had been used to capture similar costs. I could not identify any appropriate tools. I therefore, draw on the wording used in a questionnaire (Thompson and Wordsworth, 2001) that was available on the DIRUM repository which asked participants about their out of pocket expenditure. I felt the wording of the questions was appropriate as they asked whether any additional costs were incurred and if so, what the purpose of the cost was as well as the estimated amount spent.

3.3.6.3. Item 6.3. Health and social care costs

These can be captured at baseline and at the same follow up time points as the effectiveness evaluation. An adapted version of the Client Service Receipt Inventory (CSRI) (Beecham and Knapp, 1999) is presented in Appendix B.1. If data on long-term treatment costs for NCD is available in the literature, this can be included. This data should be taken from published studies.

Reflection on item 6.3

Authors from the studies included in my systematic review (Chapter 2) reported various methods for assessing the short-term (immediate) health and social care costs, these included: participant diaries, self-report questionnaires or GP medical records (see Table 1 in Appendix B.1). In order to help me decide which method would be best I consulted the wider health economic literature. I identified a Delphi study which had asked health economists about their preferred methods for capturing healthcare utilisation (van Lier et al., 2017). The authors found there is disagreement amongst health economists on whether patient-based reporting (e.g. diaries or questionnaires) or the use of secondary-level (e.g. routine medical records) data is preferred. The paper discussed the pros and cons of both methods. For instance, self-reported data is subjective and relies on the participant's accuracy and ability to recall their healthcare use, however it might be easier and cheaper to collect this data. On the contrary electronic medicals records can sometimes be incomplete, costly and typically data management systems vary across agencies making it difficult to compare similar data variables (Hughes et al., 2016).

One study (Edwards et al., 2013b) from my systematic review referenced the self-report questionnaire they had used to capture healthcare utilisation. I searched for the questionnaire on the DIRUM repository to learn more about it. The questionnaire was called the client service receipt inventory (CSRI) and was a widely validated tool that has been applied to various intervention and setting types (Beecham and Knapp, 2001, Ridyard and Hughes, 2012). The original questionnaire content relates to psychiatric services, therefore the wording and cost items included in the questionnaire were not relevant to an evaluation of a PA and SB intervention. I searched the DIRUM database and found a modified version of the CSRI by Mayer and Beecham (2005) which I draw on for the wording and structure. The specific examples I gave of health professionals were those which I had identified by analysing the content of the cost items (e.g. health professional types) included in seven good quality studies from my systematic review (see Table 1 in Appendix B.1).

3.3.6.4. Item 6.4. Participant costs

These can be captured at the same follow up data collection time points as the effectiveness evaluation. The participant cost questionnaire presented in Appendix B.1. can be used, this is an adapted version of a self-report questionnaire (Wordsworth and Thompson, 2001) retrieved from the DIRUM database. The aim of this questionnaire is to ask participants to report on their time spent taking part and travelling to the interventions of interest, the travel costs and any out-of-pocket expenses.

Reflection on item 6.4.

Out of pocket costs for PA and SB

Through the methods reported in the studies in my systematic review (Chapter 2), I found that there appears to be two main approaches to measuring participant out-of-pocket

costs and time costs: participant recall diaries or self-reported questionnaire. No studies in my review reported the measurement tool they had used to capture patient costs. I searched the DIRUM repository and identified the annotated patient costs questionnaire (Thompson and Wordsworth, 2001). The questionnaire wording and structure seemed appropriate. In order to tailor the questionnaire so as it referred to cost expenses associated with PA and SB interventions, I analysed the content of the costing methods used in the studies from my review (see Table 1 in Appendix B.1). There was a range of participant cost items measured, these included: clothes and shoes, memberships and classes fees, childcare, travel purchases (petrol based on distance travelled, public transport fee) and sports/ exercise equipment.

Recall period

In order to decide how often participants should be asked to recall their participant costs, I looked at recall periods reported in the studies in the systematic review. Overall, there was no trend in the recall period for both the diaries and questionnaire methods, with the recall period varying from 1 to 12 months. The literature on trial-based economic evaluations recommends that the recall period should align with the data collection points of the clinical effectiveness protocol (Glick et al., 2014). This option made sense since there are no existing guidelines on how frequently this data should be collected; and this option reduces participant and research burden.

Participant's acceptability

In addition to asking the participant to record their out-of-pocket costs, I felt it was important to ask participants about their preferences and their willingness to pay for a PA and SB intervention. This type of additional exploratory analysis was done by one of the studies I identified in my review and seemed to illustrate how preferences and acceptability may differ by equity subgroups (e.g. socioeconomic status) (Edwards et al., 2013b) which is an important consideration in the field of public health (Marmot et al., 2010). Furthermore, I was aware that understanding participants preferences is an important field of study within the behavioural science literature, as the acceptability of an intervention to the participants can impact on the success of the intervention (Michie et al., 2014).

Time costs and loss in earnings from the participant's perspective

I had identified two studies (Iliffe et al., 2014a, Isaacs et al., 2007) in my review, which had looked at participant costs in terms of the time the participants gave up to participate in the intervention. More specifically, they had asked participants to clarify whether they had participated during their work or non-work time, and whether the participant experienced a loss in earnings as a result. Similarly, a third study (Elley et al., 2011) from my review also asked about work, but with the intention of capturing whether the

intervention had reduced or increased sickness- and accident-related absenteeism. I therefore felt it was important to measure productivity and so incorporated key questions about this (see participants cost questionnaire in Appendix B.1).

3.3.6.5. Item 6.5. Productivity loss from the employer's perspective

If the intervention of interest is delivered in the workplace, it is recommended that absenteeism and presenteeism can also be measured using validated self-report questionnaires. The workplace limitations questionnaire (WLQ) (Lerner et al. 2001) is recommended since it one of the three main validated tools used to capture presenteeism (Kigozi et al., 2017). In addition, if the intervention is delivered in a workplace setting, the time taken away from productive work to take part in the intervention can be recorded by the staff delivering the intervention.

Reflection on item 6.5

Although no studies in my review included presenteeism in their economic evaluation, I felt this productivity measure was important to capture. This is because the economic literature argues that presenteeism is a greater contributor to employer-related productivity losses than absenteeism (Schultz et al., 2009) and that national guidelines should emphasise the importance of including presenteeism in analyses conducted from a broader perspective (Kigozi et al., 2017). The reason for the lack of studies capturing presenteeism in economic evaluations may be due to national guidelines often stating that absenteeism is preferred over presenteeism (Knies et al., 2010).

3.3.7. Item 7. How and when can effects be measured?

It is recommended that you consult your country's guidelines (see item 5) to see if there is a preferred approach to measuring the primary outcome. If you choose to measure QALYs there are a number of steps involved in the calculation of QALYs. QALYs encapsulates both quantity and health-related quality of life (HRQoL) in a single value. The number of years lived, known as life years (LYs) is calculated based on whether a participant is alive or dead. These years are adjusted according to the HRQoL reported. There are a number of HRQoL measurement tools, however the EuroQol EQ-5D is the most common tool used in economic evaluations and is freely available for research purposes (Rabin et al., 2011). The latest version of the tool asks participants to rank each of the 5 health dimensions using 5 levels of severity as opposed to 3 levels. You may prefer to use this tool as it is more sensitive at detecting differences in HRQoL. It is recommended that HRQoL is measured at the same time points as the clinical effectiveness evaluation (Glick et al., 2014). EuroQol's user guide for the EQ-5D can provide further guidance on employing the EQ-5D questionnaire (Rabin, Oemar, Oppe, Janssen, & Herdman, 2011). The long-term effects of PA and SB interventions on QALYs can be done by identifying pre-existing models, which relate a reduction in NCD (e.g. T2D, CHD and stroke) with a gain in QALYs.

Reflection on item 7

Short-term effects

Prior to commencing this PhD, I had come across the terms QALYs and EQ-5D in the literature but had not realised that there were a number of steps involved in the calculation of a QALY that go beyond the measure of HRQoL using the EQ-5D measurement tool. I was surprised to learn that the EQ-5D was a generic measure for health since the tool seemed quite limited. I recommend the use of the EQ-5D as it was the most commonly used tool reported in the studies in my review.

Long-term effects

In my review I identified two models which draw on epidemiological evidence to relate levels of PA with a reduction in TD2, CHD and stroke events, and as a consequence a gain in QALYs (Anokye et al., 2012, Campbell et al., 2015b).

3.3.8. Item 8. How can costs be valued?

Unit costs for the economic evaluation should primarily come from national published sources. For participant costs, the actual price incurred by the participant was deemed appropriate. If published sources or participant reported prices are not available for a specific resource item or category (e.g. presenteeism) then the resource can be reported in its natural units rather than be assigned a monetary value.

3.3.8.1. Intervention costs

Typically, aggregate unit costs do not exist for new interventions or even for existing individual-level PA and SB interventions. Therefore, resource items recorded in the microcosting tool can be assigned an individual unit cost before being aggregated together. Published unit costs can be used to value the individual resource items. If no unit cost is available unit costs from a similar resource can be used. All sources of unit costs can be recorded along with the date the source was accessed.

3.3.8.2. Health and social care costs

For health and social care costs national standard unit costs can be used. If no unit cost is available then the unit cost of a similar resource can be used. All sources of the unit cost and date the source was accessed can be recorded.

3.3.8.3. Participant out of pocket costs

Unit costs for participant's out of pocket costs (e.g. clothing, equipment, gym membership) will be the actual prices self-reported by the participants as in the participant cost questionnaire. Time and distance travelled will be reported in their natural units.

3.3.8.4. Employer costs

For employer costs a national average earnings can be applied where participants lose time at work due to participating in the intervention.

Reflection on item 8

Internationally the health economic literature suggests the preferred approach for valuing healthcare utilisation is by assigning national published unit costs to resource quantities (van Lier et al., 2017). The studies in my review had drawn on published unit costs in order to value healthcare use. The studies from the UK helped me identify the unit cost series for health and social care which is published annually (Curtis and Burns, 2018). I was surprised that I had not come across this unit cost series during my training in SROI when I was a Research Assistant in Public Health. Another observation I made was that the authors in my review did not report the sources they used in order to assign unit costs to the intervention resources. This may be due to most studies having reported using the budget breakdown from the funding application of the trial, rather a conduct a microcosting exercise.

3.3.9. Item 9. How can effects be valued?

Effects only need to be valued if a CUA has been performed. In addition, it is recommended that you consult your country's guidelines to identify when there is a preferred valuation approach (ISPOR, 2019). If the EQ-5D tool has been used as recommended in item 7, then EuroQol's user guide can be referred to in order to create a EQ-5D HRQoL profile for each of your participants (Rabin et al., 2011). All EQ-5D profiles already have a utility weight (also known as a preference score) assigned to them by a sample of your country's population. Assign your country's published stated utility weights to each of your participant's EQ-5D profiles. This utility weight is calculated from country-specific catalogues before being combined with data on length of life data to estimate the number of QALYs experienced over the specified time horizon. You can combine these utility weights with the time lived by the participants (during the trial) in order to estimate the number of QALYs gained or lost during the trial. Ideally, HRQoL should be collected on a schedule and you can interpolate between points to calculate the QALYs as the area under the curve. To determine cost effectiveness, many countries have guidelines to see what the maximum amount of money your country is willing-to-pay per gain in QALY. For modelling studies the unit cost/ price from the initial treatment of a stroke or CHD event, and the ongoing annual treatment for treating stroke, T2D or CHD can be taken from the existing literature. For the CCA, all disaggregated effects can be reported in their natural units.

participants HRQoL at baseline and all other data collection time points, you can assign a utility weight to each participant's score.

Reflection on item 9

I was surprised to learn that the key difference between the different types of economic evaluations is the methods they used to value the effects (outcomes) of interest. It also made me realise that unless a CBA was conducted then the result of a single CEA was limited to just one outcome measure. There is consensus amongst health economists that CBA is deemed theoretically superior to CEA and CUA as it can incorporate and monetise multiple outcomes (Drummond et al., 2015b). Yet in practice CBA is challenging to execute as demonstrated through the finding in my review and former reviews which

found no studies had performed a CBA or explain how they had monetised the outcomes (Weatherly et al., 2009, Alayli-Goebbels et al., 2014, Hill et al., 2017). During my time as a Public Health Research Assistant, I became increasingly aware that SROI was similar to CBA as it included and monetised non-health outcomes. In the SROI evaluations I had conducted, it was standard practice to assign a 'shadow' market price to outcomes which did not have a national published unit cost. I explored why the SROI methodology had not been adopted by the health economic community. Some economic analysts claimed that the SROI methodology had several theoretical problems which would need addressing before further adoption of the methodology (Fujiwara, 2015). In order to explore this further, in 2016 at the start of my PhD, I emailed a senior academic health economist from one of the health economic groups in the UK, to seek their opinion on the role of SROI and CBA in public health evaluations. The health economist informed me that they were unfamiliar with the SROI methodology. In terms of CBA, they explained that they felt CBA had the potential to generate useful information if done in accordance with the principles of the Washington State Institute of Public Policy and UK's Treasury Green Book. Nonetheless, they claimed that in general they believe CBA is often not conducted well in practice and does not address distributional issues (e.g. the UK's income distribution). This pre-PhD discussion had reassured me that the difficulty in producing a well-conducted CBA in practice was likely to explain why I had come across any CBA studies in my systematic review (Chapter 2). That is to say, until further methodological development is made in the field of CBA, I recommend in my initial framework that CEA and CUA are carried out as the preferred valuation technique.

3.3.10. Item 10. How can costs and effects be discounted to a present day value?

For trial-based economic evaluations that do not extend beyond 12 months, no discounting is necessary. For those that go beyond 12 months it is advised that you check your country's guidelines (see link in item 5) to see which discount rate is recommended for costs and effects. In the absence of country-specific discount rates, 3% can be used, as employed by the World Health Organisation (WHO) (World Health Organisation, 2017).

Reflection on item 10

Through my reading in the health economic literature, I became aware that different countries specify different discount rates. For example the WHO (2003) recommends 3% as the annual discount rate for costs and effects when there is no country-specific guidelines available. In the UK, NICE state that costs and effects should be discounted at a rate of 3.5% per annum (NICE, 2014a). In my previous public health SROI evaluations, I had come across the term discounting, however I had not considered how it was calculated. I found it interesting to read about discounting in the health economic literature, where some claim that discounting is important because a phenomenon has been observed which indicates that people typically prefer to gain benefits now and incur

the costs later (this phenomenon is known as positive time preference). This makes sense when recognising that the benefits gained in the immediate future are more certain than those gained in the distant future (Drummond et al., 2015b).

3.3.11. Item 11. What summary statistics can be presented?

For the initial summary statistics for costs and effects, means and standard deviations are recommended for continuous variables. For categorical variables, proportions are recommended along with the numerator and denominator. Where one comparator group is more expensive and less effective than the other(s), then it be evident which group is the best option to invest in. However, if your results indicate that some group is more effective but also more expensive, an incremental cost-effectiveness ratio (ICER) statistic is recommened. The calculation for the ICER is presented in Figure 4. An ICER summarises the additional cost per additional unit of effect gained. More specifically, the summary statistic allows a pre-specified decision rule to be applied in order to interpret whether the gain in effect falls within a threshold in which we are happy to pay for one additional unit of effect. The ratio can be interpreted by comparing the ratio to a willingness to pay (WTP) threshold if your country has one or an incremental health opportunity cost (Woods et al. 2016). An example of a willingness-to-pay threshold is £20,000- £30,000 per QALY gained, as applied in England (NICE, 2014).

$$\text{ICER} = \frac{\text{Cost of intervention group A} - \text{Cost of intervention group B}}{\text{QALYs of intervention group A} - \text{QALYs of intervention group B}} = \text{Cost per QALY}$$

Figure 4. Incremental cost-effectiveness ratio calculation

Reflection on item 11

Mean and standard deviation

Despite cost data typically being skewed to the right with a long tail, and HRQoL data being typically right censored (as many HRQoL measurement tools have ceiling effects), the health economic literature argues that evidence on the mean is more relevant and useful to decision makers than the median. I was surprised when I first read about this in the economic literature since in clinical effectiveness literature it is standard practice to report the median instead of the mean, if data is not normally distributed. By contrast, the economic evaluation literature it is helpful to present the mean in economic evaluations, as the median runs the risk of underestimating the amount of resources that need to be budgeted for (Gray et al., 2012). Transforming skewed data is also not recommended for economic evaluations since it is the arithmetic mean that is required rather than the geometric mean (Glick et al. 2014).

Incremental cost-effectiveness ratio

I had not calculated an ICER before, however the studies in my review indicated to me that this was the most common way to report the results of an economic evaluation. I was surprised to learn that the calculation for the ICER seemed straightforward. I think it was the different terminology used which had made me expect the calculation would be more complex. In the health economic literature, the term incremental cost and effects is commonly used. Incremental analysis is more commonly known in the clinical effectiveness literature as the between-group difference in effects. A further key difference I observed between the clinical effectiveness literature and economic literature was that there is consensus in the economic evaluation literature that a meaningful difference is typically conceptualised as being the decision-makers willingness-to-pay per unit of effect. In the clinical effectiveness literature, a meaningful difference is typically conceptualised as the minimum clinical important difference based on previous evidence and does therefore not incorporate the decision-makers willingness-to-pay into account.

3.3.12. Item 12. What adjusted analyses can be performed?

It is recommended that costs and effects are adjusted for baseline imbalances in costs and HRQoL. Multiple regression is recommended to control for the baseline covariates. Adjustment is recommended as it will improve the precision of the cost and effect estimate by reducing some of the unexplained variance in the cost and effect estimates. Both unadjusted and adjusted means and standard deviations for costs and effects can be reported (Franklin, Lomas, Walker, & Young, 2019).

Reflection on item 12

Previous research stresses that when comparing QALYs of at least two intervention groups, adjustments should be made to the mean costs and effects to account for imbalances in the participants' baseline costs and HRQoL utility (Manca et al. 2005; Glick et al. 2014). I draw on the economic evaluation literature in order to advise on how baseline imbalances could be controlled for. I found a study in one of the main health economic journals (Health Economics) which claimed that parametric tests such as multiple regression which can control for some variables, have been shown to be robust to skewed economic datasets and can generate similar results to nonparametric methods e.g. the nonparametric bootstrap method (Nixon et al., 2010).

3.3.13. Item 13. What equity subgroups can be considered?

Equity is an important objective for public health interventions, not just effectiveness. It is advised that at a minimum, the following four equity subgroups are included or discussed in your study: socioeconomic status, age, sex and medical condition. Furthermore, it is recommended that all equity-related data is collected at baseline from the participants via a self-report questionnaire.

Reflection on item 13

Equity was a key topic I learnt about through my Masters in Public Health and my Research Assistant job in Public Health. I was therefore interested to learn that one of the main challenges in public health economic evaluations, is the incorporation of equity into the analysis (Weatherly et al., 2014). The four most relevant equity subgroups for PA identified from my review, include: age, sex, socio-economic status and pre-existing medical condition. In particular, one of authors from one of the studies in my review reported that their cost-effectiveness result was particularly sensitive to age and pre-existing condition demonstrating that heterogeneity and subgroup analyses are important to consider (Campbell et al., 2015b). Literature from the UK also indicates that sex and socio-economic status are important subgroups to consider for PA evaluations (Scholes, 2017).

3.3.14. Item 14. What uncertainty analyses can be performed for trial-based economic evaluations?

It is recommended that the nonparametric bootstrapping technique is used to explore stochastic uncertainty in the sample and hence any uncertainty in the ICER point estimate. The nonparametric bootstrapping simulation randomly draw cost and effect pairs from the original dataset, in order to produce 1,000 empirical-based bootstrapped ICERs (Gray et al. 2012). This technique can produce uncertainty intervals (e.g. 95% confidence intervals) around the central estimate of cost effectiveness and can produce a cost effectiveness acceptability curve (CEAC) which assesses the probability of the intervention of interest being cost-effective at various WTP thresholds (e.g at £1,000 per QALY, £10,000 per QALY; £100,000 etc). The CEAC is recommended as it will enable you to assess the probability that an intervention would be cost-effective at different levels of willingness to pay. In addition to stochastic uncertainty, it is recommended that you assess the uncertainty associated with the methodological choices made by the analysis. This can be done using an approach called one-way scenario analysis. One-way scenario analysis involves making plausible changes to the parameters input into the ICER calculation one by one, in order to assess how a change in one parameter can impact on the ICER result.

Reflection on item 14

Stochastic uncertainty

The studies in my review did not provide guidance on how to address stochastic uncertainty in the cost-effectiveness result. Furthermore, I read in the health economic literature that characterising uncertainty is a key methodological challenge that is experienced across the whole field of health economics. In order to learn what standard practice was for addressing stochastic uncertainty in economic evaluations, I attended a 3-day training course in applied cost-effectiveness analysis at the University of Oxford. The first fundamental thing I learnt on the course was that it is not possible to calculate a

standard error statistic for a ratio statistic. The course tutors with expertises in the health economics, recommended that stochastic uncertainty in the ICER statistic could be assessed through a nonparametric bootstrapping simulation.

Methodological uncertainty
One-way scenario analysis was the most common way studies in my review addressed uncertainty in the methodological choices made by the analyst. I felt confident in doing this type of uncertainty analysis since I had performed a similar uncertainty analysis in the SROI evaluations I had conducted when I was a Research Assistant in Public health.

3.3.15. Item 15. How can the results be interpreted?

It is recommended that methodological choices are reflected on in the discussion section of the economic evaluation in order to help the read interpret the results. Key methodological choices which can be reflected on include: the perspective, the trial design, the sample size, the comparators, the costs and effects included/ excluded, the measurement tools, and the equity subgroup included. n addition, heterogeneity, generalisability and transferability are concepts which may also be help you interpret your results. Heterogeneity in terms of the comparator groups including different subgroups is also an important issue to consider in economic evaluations, since heterogeneity can drive the cost-effectiveness results. Generalisability can be reflected on in terms of whether results are relevant beyond the sample and location where the interventions are set. Transferability can be reflected on in terms of whether the results are relevant beyond the country the study has been carried out in.

Reflection on item 15
A key finding from my review was that the analyst's methodological choices and assumptions make it difficult to compare cost-effectiveness results from different studies, even if they are from the same country and perspective. This finding highlighted to me that it is important to reflect on our analyst-based choices and help the end user of the results interpret our study findings. This is especially important, since I am aware that there is a shortage of health economic expertise internationally, which means it is likely that decision-makers who may be using the results may not be familiar with how the summary decision indices (e.g. the ICER point estimate) are constructed and what the result means. Furthermore, guidance for the quality assessment checklist by Drummond et al. (2015) reports that heterogeneity and generalisability are important factors to consider when interpreting results.

3.3.16. Item 16. How can trial-based economic evaluations be reported?

As economic evaluations involve numerous methodological steps, it may be helpful to consult the main economic evaluation reporting guidelines called the Consolidated Health Economic Evaluation Reporting Standards (CHEERS). More specifically, the CHEERS checklist's explanation

and elaboration document provides examples on how different sections of the economic evaluation can be written up and presented. Although these are generic guidelines, the items included will ensure you report all the necessary features of your study in order to help the reader interpret your methods, analysis and results, and consider whether they apply to their own context. Lastly, if a CCA is conducted for the secondary analysis, it might be necessary to report this in the supplementary material since the CCA constitutes a full impact inventory of costs and effects, which is likely to be lengthy.

Reflection on item 16

For my systematic review (Chapter 2) I had drawn on the reporting guidelines called PRISMA. I had been using the Equator-network to access these guidelines. I searched this to see what economic guidelines were available and discovered that in 2013, the Consolidated Health Economic Evaluation Reporting Standards (CHEERS), a set of international reporting guidelines for economic evaluations, had been published (Husereau et al., 2013). After observing heterogeneity in the methods reported across the studies in my review, I felt the use of reporting guidelines would be one way make the study reporting more systematic and improve the comparability of methods and results across studies.

Chapter 4: Application of the initial framework to the evaluation of a co-developed PA on Referral Scheme

4.1. Introduction

4.1.1. Background and rationale

Despite the abundance of evidence on the benefits of PA, evidence on the cost-effectiveness of PA programmes is less certain, specifically for exercise referral schemes (ERSs) (Owen et al., 2017). ERSs are a common intervention strategy, with there being reportedly over 600 different types of schemes across the UK (Pavey et al., 2011a). Despite early calls for schemes to focus on helping people to incorporate PA into their lifestyles, the majority of ERSs are 12-16 week programmes which focus on encouraging structured exercise (Dugdill et al., 2005). A key systematic review found that those who participate in ERSs are more likely to improve their PA levels compared to those receiving PA advice only (Campbell et al., 2015b). That said, the review authors concluded that the specific components of ERSs, which support long-term behaviour change of PA, are unknown. Evidence from a high quality RCT on the short-term effects of ERSs found that benefits include increased PA levels of PA for those with CHD as a pre-existing medical condition, as well as lower levels of anxiety and depression amongst those with mental health or mental health and CHD as pre-existing condition (Murphy et al., 2012).

Historically, ERSs have not been underpinned by evidence-based behaviour change techniques. Recognising this, NICE in England and Wales, now recommends that all future trials on ERS clearly justify the behaviour change techniques they have incorporated into their intervention (NICE, 2014b). In the same guidance, NICE also acknowledge the lack in economic evidence on ERS, and recommend that trials measure cost-effectiveness and health-related quality of life (HRQoL) alongside their effectiveness evaluation. The present study aligns to these recommendations by assessing the cost-effectiveness of a co-developed PA on referral scheme that is underpinned by the current evidence on behaviour change (Buckley et al., 2018, Buckley et al., 2019).

4.1.2. Aims

The overarching aim of this study was to apply the initial version of the framework (Chapter 3) in order to:

1. reflect on the relevance and applicability of the framework to a real-world PA trial.
2. assess the cost-effectiveness of a co-developed PA referral scheme (Co-PARS) compared to (a) an existing exercise referral scheme (usual care) and (b) a no treatment control group.

The development of the framework was an iterative process. In my reflection boxes throughout this chapter I will revisit the framework items and consider how well they have been implemented in practice. The aim will be to describe any complexity involved in the conduct of the economic evaluation with the intention of generating theory about what is likely to be a helpful approach to address this complexity.

4.2. Methods for Aim 1: Reflections

Reflections on the planned and actual application of the initial version of the framework are documented throughout the methods and results sections. The reflections aim to provide valuable insight into of the actions I took in order to develop and apply the framework in practice. Importantly, the reflections aim to provide explanations for why I believed these actions were appropriate. That is to say, the reflections are based on my experience of applying the initial version of the framework to the Co-PARS trial. The reflections offer insight into how the framework may be modified in a future refined version.

For the costing approach I performed, I documented all of my decisions in an Excel spreadsheet so as my assumptions were transparent and I kept a record of the complexity in which was involved in my costing decisions (an example of the Excel spreadsheets is provided in Appendix C.2). In addition, I also reflect on relevant literature from the different disciplines. A key method I used in order to help me interpret the applicability of the framework was informal conversations and meetings with the Co-PARs trial team, my supervisory team, researchers from the Health Economic group at Deakin University (where I did a three week study placement) and the intervention staff. Meeting with this range of researchers and stakeholders ensured I was capturing a multidisciplinary perspective (e.g. public health, health economics, behavioural science).

In particular, a key meeting I arranged was to discuss items 4-9 from my framework (identification, measurement and valuation of costs and effects). I wanted to understand how these six items were perceived from the multidisciplinary perspectives of the experts who I was working with. I arranged a one-hour consultation with my supervisory team and key members from the Co-PARs trial team. Key members from the trial team included the PhD student (BB) and the trial manager (PW) who was also one of my supervisors. Both were selected to be involved in the meeting as they had been involved in the design of the Co-PARs intervention content and the setting up of the intervention in the leisure centres. In addition, they both had expertise in physiology, public health and behavioural science. As discussed earlier in this PhD, my supervisory team also included a range of expertise (e.g. physiology, behavioural science and public health, health economics). I therefore deemed my supervisory team's involvement in the one-hour meeting as being valuable. In total six researchers (my four supervisors and two PhD students, which included myself) took part in the consultation. Prior to the meeting I circulated an agenda to inform everyone that we would focus our discussion around items 4-9 from my framework. These items all related to the data collection plan. More specifically they related to the identification, measurement and valuation of the cost and HRQoL data. The key objectives of my consultation meeting were to: (1) identify all perspectives (stakeholders) who could experience a change in cost or effects due to the Co-PARs trial; (2) discuss the feasibility of incorporating the resource use and EQ-5D-5L questionnaires in the patient questionnaire booklet as well as discuss the feasibility of capturing intervention costs through budget breakdowns and telephone interviews; (3) explain the different approaches to valuation I planned to use; and (4) clarify roles and responsibilities for the data collection process at

the key follow up time points. Overall, the meeting was a good example of multidisciplinary working.

4.3. Methods for Aim 2: Economic Evaluation

4.3.1. Trial design

The present economic evaluation was part of a larger trial (Buckley et al., under review) which took place between 2018- 2019. The trial was a quasi-experimental design comparing three groups: (1) a co-produced PA referral scheme (Co-PARS); (2) a usual care ERS; and (3) a no treatment control group. The primary outcome measure for the trial was cardiorespiratory fitness (measured as change in VO2 max score). The primary outcome measure for the present economic evaluation was quality-adjusted life years (QALYs).

Reflection on item 1.1- What components make up a well-defined study question? (costs and effects of two or more groups)

In item 1.1. of the initial version of the framework, I explained that at least two comparator groups are required for an economic study to be identified as a “full” economic evaluation by the health economic community. The Co-PARS clinical effectiveness evaluation was being set up to compare the effects of three comparator groups. “Piggy-backing” the economic evaluation on the back of the effectiveness trial would necessitate additional data to be collected to capture resource use and HRQoL. I discussed this with the trail team and we came to the conclusion that the additional collection of economic data would not involve changing the fundamental design of the trial and would also be inexpensive to do.

Reflection on item 3- What is an appropriate study design for a trial-based economic evaluation?
In item 3 of the initial framework, I recommend that a PA trial is delivered in a setting which reflect the 'real world'. By this I mean a setting where the intervention could be rolled out in practice on a larger scale. Both the Co-PARs and usual care interventions were set within a local authority leisure centre and delivered by a qualified exercise referral practitioner (ERP) which reflected what is likely to happen under 'real world' conditions. Prior to starting this PhD I had not been involved in the design or delivery of a clinical trial before. It was through the regular monthly meetings I had with the Co-PARs trial team that I became aware of the numerous trial design decisions that had been made during and immediately following the feasibility trial in order to inform the definitive trial design. I tried to understand whether these various trial design decisions were appropriate from a health economic perspective by consulting the Drummond Handbook for Economic Evaluation for Health Care Programmes.

In the Handbook I came across a tool called the pragmatic-explanatory continuum indicator summary (PRECIS) which described several domains for trial design (Thorpe et al., 2009). In turn, this reference led me to a more refined version of the PRECIS tool, called the PRECIS-2 tool which discusses 9 domains of trial design (Loudon et al., 2015). The PRECIS-2 tools is designed to help researchers understand how trial design impacts on the degree to which the trial's results can be used to inform 'real world' decision making. I recognised this was an important factor to consider. During my Masters in Public Health I had learnt about a concept called the 'implementation gap'. The 'implementation gap' refers to the problem where the results of health research are valid in the context of a 'ideal conditions' but are not fit to inform 'real world' decisions. As the purpose of economic evaluations are to inform 'real world' resource allocation decisions, then there is consensus that economic evaluations should be conducted alongside more pragmatic orientated trials as opposed to explanatory trials. I therefore felt it is important that I recommend that the framework encourages health economic researchers to reflect on the PRECIS-2 tool from the trial outset.

4.3.2. Participants and recruitment

The target population for the trial was adults (≥18 years old) who had a health-related risk factor (e.g. hypertension, non-diabetic hyperglycaemia, obesity) and/or a health condition (e.g. diabetes, cardiovascular disease, depression) that may be improved through PA. Participants with uncontrolled health-conditions and severe psychological or neurological conditions were excluded. Participants from the Co-PARS and usual care were invited to take part in the trial by the receptionists at the leisure centres. This took place when patients visited the leisure centre to book their induction (after being referred to the leisure centre by a health professional). After participants consented to having their contact details shared, one of two PhD researchers (Ben Buckley) sent the participant an information sheet before full consent was obtained. Participants in the no-treatment control were recruited via posters, electronic invitations, email communications and ClinicalTrials.gov. Ethical approval was obtained from the North West - Preston Research Ethics

Committee (NHS Health Research Authority): 18/NW/0039. Both PhD researchers (Madeleine Cochrane and Ben Buckley) attended the ethics committee interview.

4.3.3. Comparator groups

All comparator groups received a lifestyles advice booklet, which contained information on England and Wale's national guidance for PA, diet, smoking and alcohol.

4.3.3.1. Usual Care exercise referral scheme (ERS)

Usual care followed a standard ERS model, which comprised of one 1-hour induction to the leisure centre by an exercise referral practitioner (ERP). This was followed by 12 weeks free access to the swimming pool and subsidised access (£1 per visit) to the gym and group classes during off-peak hours. During the one-hour induction, the ERP devised a 12-week exercise plan appropriate for the health condition of the participant.

4.3.3.2. Co-developed PA of referral scheme (Co-PARS)

Co-PARS included the same 12 weeks free and subsidised access to the leisure centre and one 1-hour induction. In addition, Co-PARS included four 30-minute consultations with the ERP, which took place at week 4, 8, 12 and 16. The aim of the Co-PARS intervention was to achieve sustained improvement in PA by encouraging people to incorporate PA into their daily activities. Furthermore, the Co-PARS intervention draw on evidence-based behaviour change techniques such as goal-setting and self-monitoring, that were underpinned by self-determination theory (SDT) (Buckley et al., 2018).

4.3.3.3. Control

The control group received no treatment except for the lifestyles advice booklet.

Reflections on item 2- What does a comprehensive description of the comparator groups look like?

As recommended in the initial framework, the trial protocol and CONSORT flow diagram were consulted in order to describe the three comparator groups: (1) 18-week co-developed PA on referral scheme (Co-PARS); (2) 12-week usual care (defined as existing exercise referral scheme); and (3) no treatment control. The descriptions in the protocol and CONSORT flow diagram provided an initial overview of the intervention groups, however I felt that analysing the pathway in this way did not provide enough detail about the context in which the interventions were delivered and the resource quantities involved. I noted that item 6 in my framework offered a more comprehensive approach to describing the comparator groups. This repetition of items 2 and 6 indicated to me that item 2 would not be required as a separate item, but could be incorporated into item 6.

4.3.4. Type of economic evaluation

This economic evaluation compares the costs and outcomes for the three intervention arms over a 6-month time horizon (the 6 month time horizon meant discounting was not required). The primary

economic analysis was to conduct a trial-based cost-utility analysis (CUA) from a multi-agency UK public sector perspective (agencies are discussed in the subsequent section). The secondary economic analysis was to conduct a cost-consequence analysis (CCA). The methodological approach was informed by the recommended framework in Chapter 3. Full details on how the initial version of the framework has been applied to this trial is documented alongside the study reflections findings in section 5.3.4.

Reflections for item 1.2- What components make up a well-defined study question? (Primary Analysis)

In item 1.2 of the initial version of the framework, I advise that it is helpful for analysts to check their country's guidelines (referred to as the 'reference case' by the health economics community) to see if there is a preferred analysis type for that country. The UK's reference case recommends a CUA is performed for the primary base case analysis. This explains why I decided to conduct a CUA for the primary analysis. I felt it was important that my analysis technique was consistent with other UK-based economic evaluations. The UK's reference case specifies CEA (where outcomes are reported in their natural units), CBA and CCA are appropriate for additional analyses. In my systematic review (Chapter 2), I concluded that a CCA conducted alongside a CUA or CEA would be helpful for the multi-sector audience of the PA and SB interventions to see a breakdown of the costs and outcomes which relate to them. I had not identified any examples of CBA applied to PA through my review, so I did not feel confident in carrying out a CBA without accessing additional training. At the time I was planning my analysis, I did not come across any practical training opportunities in the UK.

4.3.4.1. Perspective

The CUA was conducted from the multi-agency public sector perspective, which included: primary and secondary healthcare agencies, local government (payer), the leisure centre (provider and set up costs) and the research institute (set-up costs), as recommended for interventions outside the healthcare setting (NICE, 2014a). The CCA included the public sector agencies outlined for the CUA, as well as the participants' perspective. Cost categories for each perspective and economic evaluation type are outlined in Table 6.

Table 6. Perspective, costs categories and economic analysis

Sector	Perspective	Cost category	Economic evaluation type
Public	Payer intervention costs	Intervention operating costs	CUA; CCA
			CCA
	Leisure Centre (Provider)	Intervention additional operating costs	CUA; CCA
	Healthcare sector	Primary healthcare	CUA; CCA
		Secondary healthcare	CUA; CCA

		Prescriptions	CCA
	Research Institute	Intervention set up costs	CCA
Private	Participant (Private individual)	Time costs	CCA
		Travel costs	CCA
		Out-of-pocket costs	CCA

CUA – cost utility analysis, CCA – cost consequence analysis

Reflection for item 1.2- What components make up a well-defined study question? (Perspective)

In item 1.2 of the initial version of the framework, I advise that the primary analysis is conducted from the public sector perspective. In my previous role as a Research Assistant, I had frequently been involved in identifying the stakeholders involved in the service evaluations I was performing. I believe this experience of identifying stakeholders helped me recognise that the health economic term 'perspective' is similar to the public health term 'stakeholder'. Furthermore, I was able to recognise that the payer's point of view in the Co-PARS evaluation is the local government and that the provider was the local authority leisure centre.

Prior to conducting the systematic review, I had not considered set-up costs as being different to the delivery costs. In my systematic review (Chapter 2), the study by Edwards et al. (2013) had identified set-up costs as a key cost category. That said, during my research placement at Deakin University, the health economists informed me that they do not usually include set-up costs in their economic evaluations. They explained that their country's guidelines preferred them to evaluate interventions as though they are operating as a 'steady state'. The UK's economic guidelines do not specify whether public health evaluations should include set-up costs, I therefore chose to take a more comprehensive approach and included it**.**

4.3.4.2. Data collection procedure

The trial had three key data collection points. Baseline data was collected 1-3 weeks prior the intervention start date. The two follow up time points took place 12 weeks and 6 months after the intervention start date. Baseline and 12-week data collection took place in the university laboratories. Participants had various physiological measures taken which are described in the main trial (Buckley et al. under review). The demographic and economic were measures using self-reported questionnaires. Six-month data collection was collected via post; the PhD researcher (Madeleine Cochrane) was responsible for coordinating this data collection. In addition Madeleine Cochrane was responsible for the data handling and analysis of all economic data from the three time points. Additional intervention cost data was sought by Madeleine Cochrane through a face-to-

face consultation with the gym manager and ERP at each leisure centre. The researcher Ben Buckley, who was also responsible for all clinical data, collected data at baseline and 12 weeks.

4.3.5. Cost measures

Costing necessitated the collection of two types of data: (1) quantities; and (2) unit costs. For the present study, primary data collection methods were used to estimate resource use quantities, while unit costs came from secondary sources except for participant costs, which used actual prices reported by the participants. Measurement tools used to capture resource use quantities are outlined below and described in Appendices B.1 and C.3. Research costs were not included in the study. Costs categories included are detailed in the subsequent sections.

Reflections on item 4- What costs are important and relevant?

In item 4 of the framework I provide examples of five perspectives that may be relevant to the analysis of a PA intervention: the payer, the provider, health care, the participant and the employer. In order to help me identify the organisations and individuals which related to these five categories for the Co-PARS trial I arranged a one-hour consultation with my supervisory team and key members from the Co-PARs trial (see section 4.2 for further details on the consultation). At the meeting, the trial team confirmed that the payer of the intervention was the local authority and that the leisure centre was the provider. In terms of including healthcare costs, although no-one at the meeting had experience of collecting these costs in a trial-based evaluation, there was recognition that from a public health perspective the Co-PARs intervention could have the potential to reduce the demand for primary care visits and prescribed medications. Similarly, no-one at the meeting had experience of including participant out-of-pocket costs in their trial-based questionnaire booklets, however there was consensus from a behavioural science perspective as well as an equity (public health) perspective that the financial and time costs incurred by the participant were important as they might reveal hidden barriers which had not been considered e.g. the type and amount of time given up to travel to and participant in the Co-PARs intervention. I feel the health economic perspective of time being given up by the participant was an interesting discussion which many of us at the meeting had not previous incorporated into our data collection questionnaires. The meeting also helped me recognise the benefits of meeting with the trial team to capture a multidisciplinary perspective.

4.3.5.1. Intervention costs: quantities

It was initially intended that a budget breakdown of the intervention costs would be acquired from the local government who allocate funding to the ERS programmes across the Liverpool region. This budget breakdown was not available since funding was based on a payment transfer system. Instead, a microcosting exercise was conducted. The CONSORT flow diagram from the larger trial (Appendix C.1) was firstly used to identify key intervention activities for each site. Further detail on the activities (e.g. how they were delivered and/or modified in practice) were captured

retrospectively through a one-hour consultation with the ERP and leisure centre manager at each intervention site. The TIDieR framework checklist was used to guide the discussion. This data was recorded in a modified version of the microcosting tool (Appendix C.1 provides an example on how the microcosting tool was applied).

Reflections on item 6.1- How can costs be measured?

Intervention operating and set up costs (planned and additional costs)

It was not possible to access a budget breakdown for the ERS from the local authority since the payment system for the ERS programme was a block contract. As a result all intervention resource data were collected retrospectively through face-to-face meetings with the leisure centre managers (n=2), ERPs (n=2) and the trial manager (n=1). Research costs were not included in the analysis, however the trial manager was consulted as they were able to recall the number of meetings that were needed between the ERP, leisure centre managers and local authority staff in order to set up and provide information about the new Co-PARS intervention. I had not been involved in the development and set-up of the Co-PARS intervention at the leisure centres since this preliminary work began at the beginning of 2016, before I had begun my PhD project. Therefore I was not able to ask the trial team to document these set up costs. I feel that a key learning point for me was that it would be helpful to have a health economic perspective involved from the inception of the project. In addition, another key learning point for me was about the benefits of being flexible with the intervention staff. I had originally planned to arrange telephone interviews with the intervention staff to collect intervention resource use data. Nonetheless, the leisure centre staff requested if they could meet face-to-face and so I changed my approach in order to ensure this resource use data was collected.

I also changed my microcosting approach. I had originally recommended that the CONSORT flow diagram and trial protocol could be used to identify resource use quantities. However, I found that this approach did not comprehensively identify all intervention resources. Prior to collecting data on the intervention costs, I had become aware through informal conversations with the trial team that the intervention was more complex than I thought and involved multiple components. I therefore became aware that the brief telephone interview schedule (Appendix B.1) and the simple microcosting tool (Table 5 in Chapter 3) I had proposed in the original framework in Chapter 3 were unlikely to be able to capture the complexity of the Co-PARs and usual care intervention. I sought advice around capturing the complexity of the intervention from one of the senior health economists from Deakin University who I had met during my study placement and who had experience of costing public health interventions. They informed me about an approach called ‘pathway analysis’ which was used in the ACE-Prevention project (Vos et al., 2007). ACE-Prevention was a national project in Australia involving Deakin University, which assessed the cost-effectiveness of preventative interventions. I was informed that pathway analysis can be helpful for conceptualising a complex intervention as it asks: “*who does what, to whom, when, where, and how often*?” (Vos et al. 2007: 11). I recognised that this analysis framework was very similar to item 2 of the critical appraisal checklist by Drummond et al. (2015) which I had used in my systematic review (Chapter 2). The critical appraisal checklist explains that identification of intervention costs requires information about: “*who does what to whom, where, and how often*?” (Drummond et al. 2015: 45). This indicated to me that there was some evidence of consensus across the health economic literature, on how interventions can be

conceptualised. I made a similar observation with the template for intervention description and replication (TIDieR) framework which is published in the public health and behavioural science literature. Similar to the ACE-prevention and Drummond's checklist, the information deemed important for understanding an intervention using the TIDieR checklist is: *"why, what, who provided, how, where, when and how much, tailoring, modifications and how well?".* The TIDieR checklist includes additional factors such as 'how', 'tailoring', 'modifications' and 'how well'. The latter two concepts (modifications and how well) refer to the difference between the planned intervention and the intervention that was actually delivered in practice. I recognised that these two items had the potential to capture data which distinguished between the planned costs and actual costs (e.g. any additional costs) which I had identified in one of the studies in my systematic review (Edwards et al., 2013b).

From the perspective of the behavioural science and public health literature, the authors of the TIDieR checklist explain that a comprehensive description is important for replicating the intervention and using the results as historically interventions have been poorly described in clinical effectiveness literature (Hoffmann et al., 2014). Poor reporting is thought to be due to journal authors and editors not having provided guidance on what to report (Schroter, Glasziou, & Heneghan, 2012). From a health economic perspective, a comprehensive description of the intervention is important as the resources used to deliver the interventions may drive the cost and explain the cost-effectiveness result (Anderson, 2010). From the perspective of the behavioural science literature, the description is important and may explain some of the mechanisms of behaviour change which can help with the design of a more effective and efficient intervention (Michie et al., 2013). Overall, these observations helped me identify an area for where multidisciplinary working had the potential to succeed.

4.3.5.2. Intervention costs: unit costs

Published unit costs were used as recommended in the framework (Chapter 3). This included costs for staff time to deliver and set up the intervention (e.g. time of the ERP, Receptionist and Researcher). The salary unit costs used included overhead costs in their calculations, therefore the cost of the capital equipment used to deliver the intervention (e.g. the private room, IT system and telephone in the leisure centre and research institute) was not included in the analysis, but these items are still quantified in their natural units. Printing costs derived from published unit costs, while a proxy cost was assigned to the subsidised membership since no unit cost was available for this item. All unit cost calculations for the microcosting exercise are provided in Appendix C.2.

4.3.5.3. Healthcare resource use: quantities

An adapted version of a widely used healthcare utilisation questionnaire called the client service receipt inventory (CSRI) (Mayer and Beecham, 2005, Beecham and Knapp, 2001) was completed by participants. The questionnaire asked participants to recall their healthcare use over the

previous 6 months, therefore participants were only asked to complete the questionnaire at baseline and the 6-month data collection point.

4.3.5.4. Healthcare resource use: unit costs

Healthcare unit costs were sourced from the England's annual Health and Social Care unit cost publication (Curtis and Burns, 2018). Participants were not asked to distinguish between their hospital outpatient and day case visits, consequently an average of the unit cost for outpatient and day case visits was assigned from England's database for secondary care reference costs (NHS Improvement, 2018). Medication costs were reported in their natural units. Cost calculations and unit cost sources are provided in Appendix C.2.

4.3.5.5. Participant resource use and opportunity cost of time

An adapted version of the annotated patient costs questionnaire (Thompson and Wordsworth, 2001) was completed by the participants at 12-weeks and 6 months (Appendix C.3). In brief, the questionnaire captured participants' out-of-pocket costs and any time they spent participating in the intervention including whether this was time lost in work. In the CSRI healthcare questionnaire (described in the previous section), participants were asked to tick whether they paid privately for any healthcare costs. As the study was from the public sector perspective, these costs were reported in the CCA but excluded from the CUA.

4.3.5.6. Participant costs: unit costs

Time was reported in natural units (hours/ minutes). Out-of-pocket costs for equipment and leisure centre memberships were reported and valued using the actual prices reported by the participants. Private healthcare utilisation was reported using unit costs from the UK's annual Health and Social Care unit cost publication (Curtis and Burns, 2018).

4.3.5.7. Currency, price year and conversion for all costs

Dates of all prices are reported in Appendix C.2. Nearly all unit costs came from secondary sources for the current price year (2018/19). Where the price year differed, a price-year adjusted cost estimate was calculated by adjusting unit costs to the target year (2018/19) and by applying the UK's GDP deflator index (HM Treasury, 2019). All unit costs came from UK sources meaning it was not necessary to convert currencies.

Reflection on item 1.4- What components make up a well-defined study question? (Time Horizon)

In item 1.4 of the framework presented in Chapter 3, I advise that if there are sufficient data, time and expertise, then a decision model can be conducted after the trial-based economic evaluation. There were two main barriers which precluded a decision-model being built for my analysis of the Co-PARS trial. These related to the time it would take to build a decision model from scratch and secondly my lack of my involvement in the early stages of the trial.

Firstly, through discussions with other health economists during my study placement with Deakin Health Economics group we concluded that building a decision model from scratch for the Co-PARs economic evaluation would be a time-consuming process and would be beyond the scope of the aims of my PhD project. Therefore, rather than building a model from scratch we concluded it would be sensible to contact the authors of the pre-existing models I identified in my systematic review (Chapter 2) to see if they were able to share the data they had used to build their model. I contacted the two key corresponding authors of the PA models from my systematic review twice, however I received no reply from either author.

Secondly, another barrier related to the fact that I had conducted my systematic review at the same time in which the feasibility study for the Co-PARS trial was being carried out. I therefore conducted my review before the study design for the definitive trial of Co-PARS had been finalised. I was aware that PA levels was being measured as the primary outcome for the feasibility study. For this reason, I limited the eligibility criteria of the studies in my review to the assessment of PA levels only. Findings from the feasibility study, led the trial team to decide that there was not be enough time or resource to assess PA levels as the primary outcome for the definitive clinical effectiveness trial (since a large sample size would be required), therefore the trial was designed to be powered to collect VO2 max scores as the primary outcome. I had not identified any decision models in my systematic review where short-term VO2 max or EQ-5D scores had been used to link these short-term effects with long term impact (e.g. increase in QALYs and reduction in treatment costs).

Both these barriers highlighted to me the need for sufficient time and appropriate data in order to build a decision model. This first-hand experience helped me understand why most analysts in my review had not gone on to produce a decision model after conducting a trial-based economic evaluation. This experience also highlighted to me the importance of the health economic analyst becoming involved in the feasibility stage of the trial rather than the definitive stage. I feel I would have benefit from getting involved earlier so as I could consider how the data completeness results might impact on the economic analysis plan for the definitive trial.

Reflections on item 10- How can costs and effects be discounted to a present day value?
As I did not include future costs and effects in the analysis I was not required to apply a discount rate. The trial-based economic evaluation assessed costs and effects over a 6-month time horizon only.

4.3.6. Economic outcome measures

4.3.6.1. Quality-adjusted life years

EuroQol's validated and widely used generic measurement tool called the EQ-5D-5L (EuroQol 5 dimension, 5 level) measured HRQoL (Rabin et al., 2011). Pre-existing preference weights (HRQoL index scores) for the UK population were matched to each health state to calculate each participant's HRQoL utility score. This is the preferred method for the main economic analysis conducted from the UK's public sector perspective (NICE, 2014a). The final step to deriving the EQ-5D scores involved mapping the EQ-5D-5L index scores to the EQ-5D-3L using a recommended mapping function (van Hout et al., 2012). Mapping is recommended in NICE's recent position statement for the valuation of the EQ-5D-5L measurement tool (NICE, 2018). The EQ-5D-5L was completed at baseline, 12 weeks and 6 months. In order to calculate QALYs for each participant, an average of each participant's three EQ-5D scores was used and then combined (through multiplication) with length of life, which at 6 months was 0.5 life years.

Reflections on item 5- What effects are important and relevant?
As recommended in item 5 of the framework, I used a single generic measure of health benefit for the primary outcome. Overall, item 5 was straightforward to apply. I feel this was because there is clear consensus on what the preferred methodological approach is for studies conducted from the UK. For instance, the UK reference case clearly stated that the Quality-Adjusted Life Year (QALY) is the preferred measure for the primary outcome of public health economic evaluations evaluation adult-based interventions (NICE, 2014a). More specifically, the EQ-5D measurement tool is recommended to measure the HRQoL part of the QALY calculation. In order to provide a breakdown of the QALY calculation I presented this in the CCA which was straightforward to do.

4.3.6.2. Willingness-to-pay preferences

At 6 months, participants across all three groups were asked about their willingness-to-pay (WTP) for a hypothetical PA on referral scheme that involved one-to-one consultations with an ERP and access to leisure centre services. WTP questions are outlined in Appendix C.3.

4.3.7. Analysis

4.3.7.1. Complete case analysis

The CUA was a complete case analysis, which aligned to the intention-to-treat (ITT) principle. For multi-item measurement tools where only a small proportion of the data was missing (less than 10%) it was deemed acceptable to impute the mean of each group for participants missing an item

(Eekhout et al., 2014). In the CCA medication data, participant costs and WTP questions were reported based on the number of available-cases. Heterogeneity between the groups was assessed through descriptive statistics by comparing the groups' baseline characteristics.

4.3.7.2. Summary statistics

A patient-level analysis was performed, where costs and QALYs for each participant were presented. Total mean costs were calculated using the absolute intervention and healthcare costs incurred between baseline and the 6-month follow up period. Area under the care data for the period between baseline, 12 weeks and 6 months estimated the change in QALYs. From here, measures of central tendency for the QALYs and costs were calculated for each group. Co-PARS mean costs and QALYs were compared to the mean values for: (1) usual care; and (2) no-treatment control. The incremental cost-effectiveness ratio (ICER) was estimated. Measures of sampling variability are presented alongside the point estimates (e.g. standard deviations and 95% confidence intervals). Results of both unadjusted and adjusted analyses are presented. For the adjusted analyses, multiple regression was performed to adjust for baseline differences in HRQoL utility values and healthcare costs.

Reflections on item 12- What adjusted analyses can be performed?
The six month costs and QALYs were adjusted for baseline imbalances in healthcare costs and HRQoL scores using multiple regression. Adjusting for baseline imbalances using multiple regression is a widely used statistical approach within the clinical effectiveness literature. I therefore identified several practical examples on how to conduct this analysis in practice. I therefore felt confident my approach and results (Table 8).

4.3.7.3. Uncertainty analysis

Since it is not possible to estimate standard error for a ratio statistic (Gray et al., 2012) stochastic uncertainty was assessed through a bootstrapping simulation of 1,000 bootstrapped replicates. From here, cost-effectiveness planes and cost-effectiveness acceptability curves were produced to help show the uncertainty in the summary statistics. In addition, a one-way scenario analysis was performed to consider the variation and uncertainty in the total cost estimate when a different unit cost was used for outpatient and day case patient.

4.3.7.4. Equity considerations

At baseline participants self-reported equity-relevant demographic data: socio-economic status (postcode area), age, sex, medical condition referred for, number of medical conditions, ethnicity and occupation status. Epidemiological evidence from the UK highlights that the following subgroups were more likely to be physically inactive: females, aged 55 or over, obese, living in the most deprived quintile (Scholes, 2017). The equity impact analysis was exploratory as it was performed for participants within the Co-PARS group only. This was because only the Co-PARS group had a large enough sample, to have subsamples of 10 or more observations.

4.3. Results of the Economic evaluation

4.3.1. Baseline characteristics

A total of 68 participants were enrolled onto the trial between March- August 2018, 55 of whom provided measures at all three time-points (Figure 5). The largest group across all three time points was the Co-PARS group. Participant characteristics for each comparator group are presented in Table 7. The table shows that the proportion of individuals in each group was similar for ethnicity, sex, age, referral reason and co-morbidity status (having more than one health condition). In general, across all groups the majority of participants were white British, had a co-morbidity and were aged 50 years or over. Almost half (48%, n=12/ 25) of the Co-PARS group lived in an area classed as England's most deprived quintile compared to just over a quarter (28.5%, n=4/14) of the control group, and a fifth (18.8%, n=3/ 16) of the usual care group. Control group participants were more likely to be in full or part time employment than both Co-PARS and usual care participants. Furthermore, the Co-PARS and usual care groups included participants who were absent from work due to long-term sickness/ disability or retirement, while the control group included no participants with these characteristics.

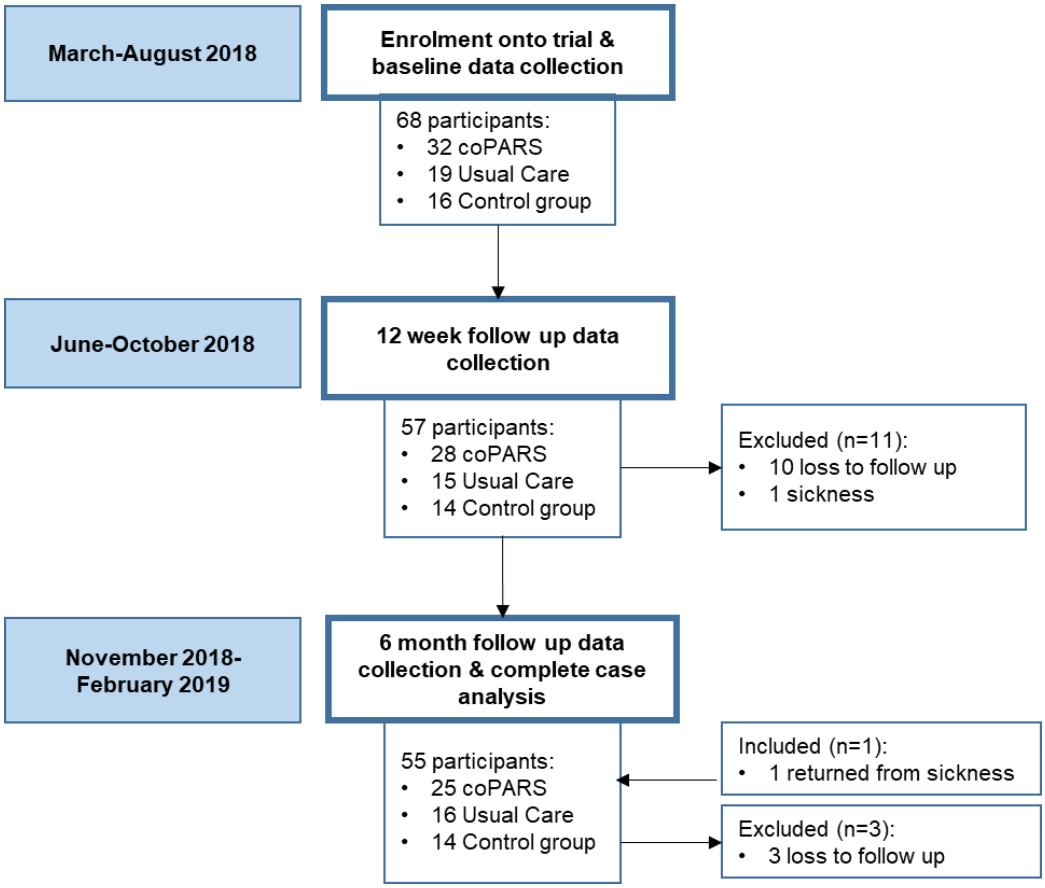


Figure 5. Flow diagram of participants enrolled on Co-PARS trial

Table 7. Baseline Characteristics

Characteristic		Comparator groups	
	Co-PARS (n=25)	Usual Care (n=16)	Control (n=14)
EQ-5D score	0.640 (0.039)	0.724 (0.049)	0.872 (0.053)
Live in top 20% most deprived area nationally	48.0% (n=12)	18.8% (n=3)	28.5% (n=4)
Ethnicity: White British	84.0% (n=21)	93.8% (n=15)	78.6% (n=11)
Occupation: Full-time employment	20.0% (n=5)	25.0% (n=4)	71.4% (n=10)
Occupation: Part-time employment	4.0% (n=1)	25.0% (n=4)	21.4% (n=3)
Occupation: Retired	24.0% (n=6)	25.0% (n=4)	0.0% (n=0)
Occupation: Long-term sickness or disability	28.0% (n=7)	12.5% (n=2)	0.0% (n=)
Sex: Female	60% (n=15)	56.3% (n=9)	57.1% (n=)
Main referral reason: Cardiometabolic	60.0% (n=15)	43.8% (n=7)	64.3% (n=)
Main referral reason: Mental Health	24.0% (n=6)	18.8% (n=3)	21.4% (n=)
Main referral reason: Musculoskeletal issues	12.0% (n=3)	31.3% (n=5)	7.1% (n=1)
Co-morbidity	88.0% (n=22)	100% (n= 16)	78.6% (n=11)
Mean age (years)	55.9(±13.7)	55.3(±16.3)	49.6(±17.3)
Aged 55 and over	56.0% (n=14)	56.3% (n=9)	50.0% (n=7)

4.3.2. Cost-utility analysis

4.3.2.1. Summary statistics

Table 8 reports the CUA results adjusted for baseline differences. At 6 months follow-up mean incremental QALYs were higher in the Co-PARS group compared to usual care (+0.021, 95% CI: -0.008 to 0.05) and the control group (+0.003, 95% CI: -0.03 to 0.036) (Table 8). At 6 months mean incremental costs were higher in the Co-PARS group compared to usual care (+£322.34, 95% CI: £-476.53 to £1,121.20) and the control group (+£471.27, 95% CI: £-363.95 to £1,306.48). In summary, the Co-PARS group costed more but gained more QALYs at 6 months follow up. Nevertheless, the 95% CIs indicate that there is uncertainty in whether the true difference is

negative or positive. In terms of cost-effectiveness, the point estimate of the ICER for Co-PARS compared to usual care was £15,349 per QALY. Employing NICE's WTP threshold suggests Co-PARS is cost-effective compared to usual care. By contrast, the ICER for Co-PARS vs. control group was £157,088 per QALY. Using NICE's WTP threshold suggests Co-PARS is not cost-effective compared to the control group. Results of the CUA based on the original unadjusted data are presented in Appendix C.4. See Table 9 in the CCA for breakdown in the results.

Table 8. Results for CUA at 6 months

Variable	Co-PARS Mean at 6 months (SE) per participant	Usual Care ERS Mean at 6 months (SE) per participant	Control Mean at 6 months (SE) per participant
	QALYs & Costs at 6 months*		
QALYs	0.385 (SE:0.008; 95% CI: 0.37 to 0.40)	0.364 (SE: 0.009; 95% CI: 0.35 to 0.38)	0.382 (SE: 0.010; 95% CI: 0.36 to 0.40)
Total costs	£852.82 (SE: £201.70; 95% CI: £447.90 to £1257.75)	£530.49 (SE:£252.05; 95% CI: £24.27 to £1,036.51)	£381.56 (SE: £270.01; 95% CI: £-160.51 to £923.63)
	Incremental QALYs & Costs *		
Incremental QALYs: Co-PARS vs Usual Care	0.021 (SE:0.012; 95% CI: -0.008 to 0.05); p-value=0.230		
Incremental QALYs: Co-PARS vs Control	0.003 (SE:0.013; 95% CI: -0.03 to 0.036); p-value=1.000		
Incremental Costs: Co-PARS vs Usual Care	£322.34 (SE:£322.71; 95% CI: £-476.53 to £1,121.20); p-value=0.968		
Incremental Costs: Co-PARS vs Control	£471.27 (SE:£337.39; 95% CI: £-363.95 to £1,306.48); p-value=0.506		
	ICER statistic at 6 months*		
Variables	Co-PARS vs Usual Care	Co-PARS vs Control	
ICER point estimate	£15,349 per QALY	£157,089 per QALY	

	ICER 95% CIs at 6 months**	
Variables	Co-PARS vs Usual Care	Co-PARS vs Control
95% CI for ICER based on 1,000 bootstrapped simulations	-£188,650 to £229,599	-£16,035 to £374

*Adjusted for baseline imbalances in EQ-5D score and healthcare costs; ** original unadjusted data

Reflections on item 1.2- CUA as the primary analysis

The initial framework advised the following analysis items which relate to the analysis approach: conduct a CEA (or CUA) as the primary analysis (item 1); present an incremental analysis (item 11); adjust for baseline imbalances (items 12); and assess uncertainty in the results (item 14). I conducted a CUA in line with the UK's reference case. The first part of the CUA calculation was similar to the methods I had come across for public health clinical effectiveness evaluations. I observed a small difference in terminology between the effectiveness and economic evaluation literatures. In effectiveness evaluations the analysis is typically referred to as 'the between group difference in effects'. In economic evaluations, the same analysis is referred to as 'the incremental effects' and includes an assessment of the 'the incremental costs' (Table 8). Economic evaluations also have an additional second part to the 'between group difference'/ 'incremental analysis' which does not feature in effectiveness evaluations. The between group difference in costs is typically compared to the between group difference in effects so as the 'cost per effect' (e.g. cost per QALY) can be presented. I felt confident in how to calculate the incremental cost-effectiveness ratio, as I found a plethora of examples in the health economic literature. Nevertheless, the confidence intervals for the cost data in Table 8 indicate there is a large amount of sampling variation and thus a considerable amount of uncertainty in the results. This indicated to me that it may have been inappropriate to perform a CUA on the sample from this trial as the small sample size and non-randomised nature of the trial is likely to have influenced the results for the within and between group analyses.

Reflections on item 11- What summary statistics can be presented?

An important observation I noted was that when I informed the non-health economic researchers from the Research Institute I was studying at, that the Co-PARS vs usual care result was "£15,349 per QALY" they initially interpreted this result to mean that the Co-PARS intervention was decisively cost-effective. This made me aware of the consequence of reporting the ICER result by itself. I felt ICER statistic by itself did not inform people about the uncertainty that underpins my data and thus the ICER result. Similarly, through the informal conversations I had with non-health economic researchers at a PA conference in London in 2018 (the International Society for PA and Health congress), I became aware that many of the non-health economic researchers I spoke to had heard of the cost-effective result term the 'ICER', but they had not heard of the uncertainty analyses that health economists typically present alongside ICER results e.g. cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC). I therefore feel that item 11 (presentation of the ICER) could be combined with item 14 (presentation of the uncertainty analyses) in order to encourage these results and analyses to be reported and interpreted in tandem.

4.3.2.2. Uncertainty analyses

The ICERs' 95% confidence intervals generated through the non-parametric bootstrapping simulation confirm the findings of the adjusted incremental analyses whereby there is substantial uncertainty in whether Co-PARS is associated with lower or higher QALYs and costs, and thus whether Co-PARS is likely to be cost-effective (Table 8).

Cost-effectiveness planes

The scatter plot of the bootstrapped incremental costs and QALYs comparing Co-PARS to usual care (Figure 6) shows that there is substantial uncertainty in whether Co-PARS generates a change in QALYs and costs compared to usual care. This is evident since the bootstrapped ICERs fall across all four quadrants of the cost-effectiveness plane. By contrast, the scatter plot of the bootstrapped incremental costs and QALYs comparing Co-PARS to the control group (Figure 7) shows that it is likely the Co-PAR group generates higher costs and less QALYs than the control group. This is evident since the majority of bootstrapped ICERs fall on the north-west quadrant of the cost-effectiveness plane.

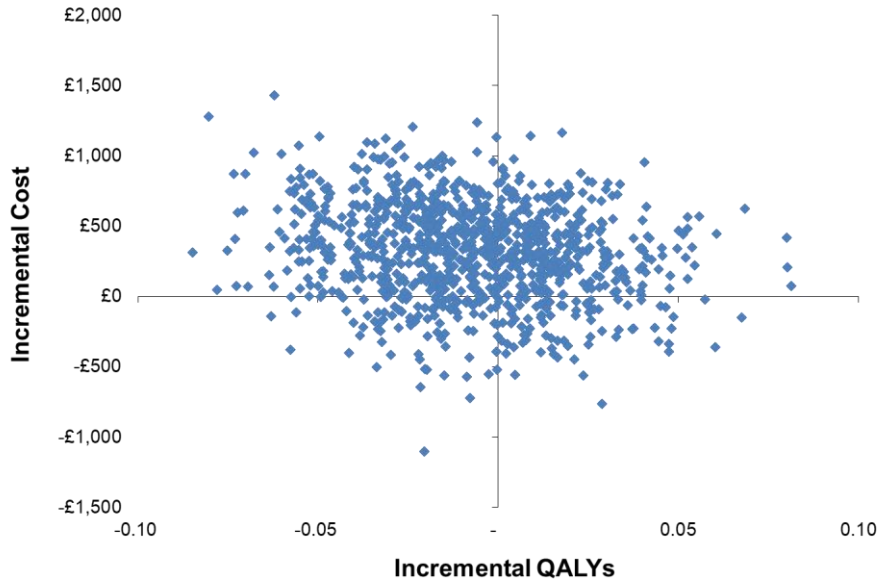


Figure 6. Cost-effectiveness plane for Co-PARS vs usual care at 6 months

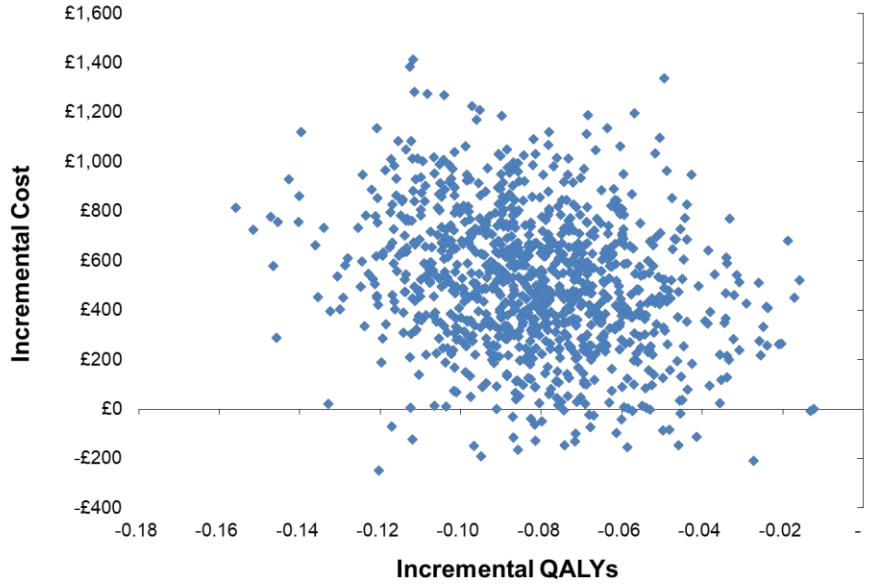


Figure 7. Cost-effectiveness plane for Co-PARS vs control at 6 months

Cost-effectiveness acceptability curves

Cost-effectiveness acceptability curves (CEAC) have been presented to help show how the decision on whether the probabilistic findings are deemed cost-effective depends on NICE's maximum willingness to pay threshold (£20,000-£30,000 per QALY). As shown in Figure 8, at a threshold of £30,000 per QALY, there is a 26.8% probability that Co-PARS will be cost-effective compared to usual care. This probability increases slightly when the threshold increases, indicating that the results are influenced by the difference in costs, rather than QALYs. Figure 9 shows that even at a threshold of £50,000 per QALY, Co-PARS had zero chance of being cost-effective

compared to the control group. This implies that the results are being driven by the difference in QALYs.

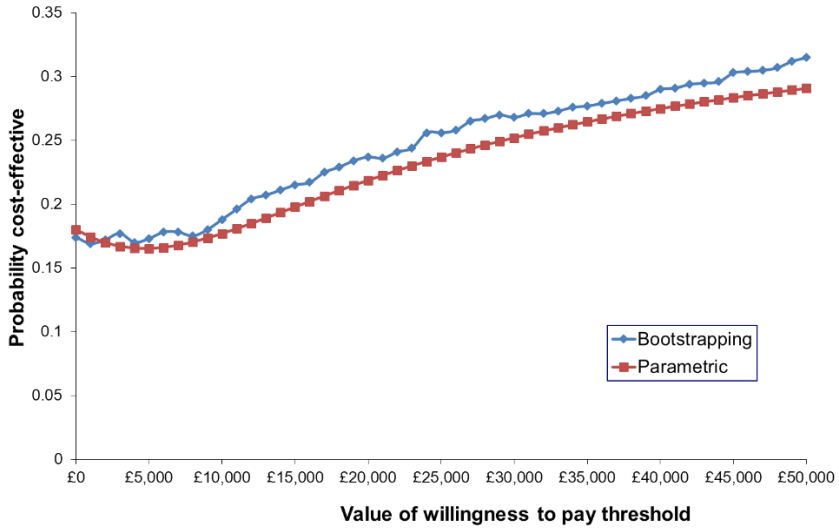


Figure 8.Cost-effectiveness acceptability curve showing the probability of short-term (at 6 months) cost-effectiveness for the Co-PARS group vs usual care group at different willingness to pay per quality adjusted life year (QALY) thresholds

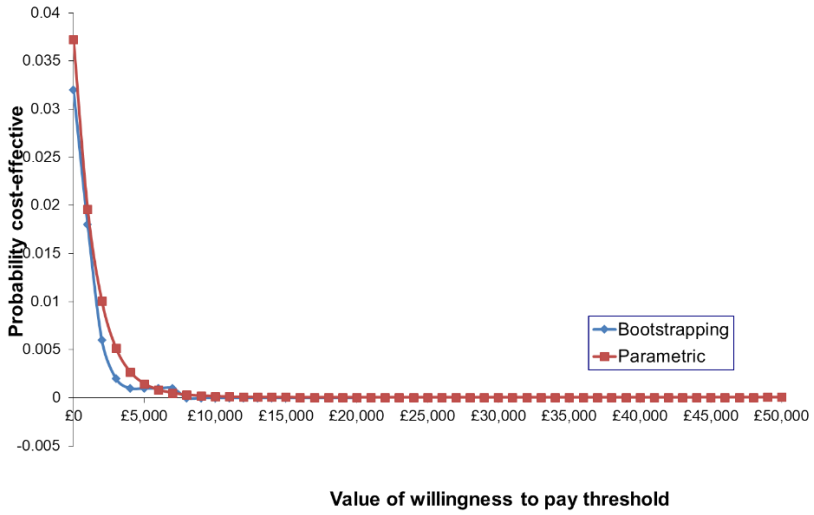


Figure 9. Cost-effectiveness acceptability curve showing the probability of short-term (at 6 months) cost-effectiveness for the Co-PARS group vs control group at different willingness to pay per quality adjusted life year (QALY) thresholds

Reflection for item 14- What uncertainty analyses can be performed for trial-based economic evaluations?

Prior to conducting this PhD project, in the public health evaluations I had conducted, I had primarily characterised the uncertainty in my results by using hypothesis testing. This was because the purpose of the public health research I had been involved in was about making inferences about a particular phenomenon. Through this PhD project and my reading of the health economic literature, I became aware that the primary purpose of economic evaluations is not to test hypotheses and make inferences about phenomenon, but to inform unavoidable decisions about resource allocation. There is consensus amongst leading health economists in the UK that estimation is more appropriate than hypothesis testing when assessing uncertainty in the results of economic evaluations (Drummond et al., 2015a).

The first type of uncertainty measures I presented using an estimation approach were the 95% confidence intervals (CIs) for the incremental QALYs and costs (Table 8). These 95% CIs are the uncertainty measures that are commonly reported alongside the results of effectiveness evaluations and so although I had tended to report only p-values as a Research Assistant in Public Health, I was familiar with the concept of 95% CIs. The second type of estimation approach I used for characterising uncertainty in my cost-effectiveness result (the ICER) was the non-parametric bootstrapping method. This method enabled me to build an empirical estimate of the sampling distribution of the ICER. This method is standard practice in applied economic evaluations (Gray et al., 2012). The estimated sampling distribution enabled me to visually present the stochastic uncertainty associated with the ICER result through the presentation of cost-effectiveness planes and cost-effectiveness acceptability curves (CEACs).

Cost-effectiveness planes and CEACs are unique to the field of health economics. Prior to this PhD, I had not come across these types of graphs in the public health effectiveness literature. Understanding how to generate an empirical sampling distribution for the ICER and go on to present cost-effectiveness planes and CEACs involved specialist knowledge. I was unable to find step-by-step practical guidance on how to this and so it was only through a specialist three-day Applied Methods in Cost-Effectiveness Analysis course at the University of Oxford that I gained this practical knowledge. This made me realise that it would be challenging to incorporate this knowledge in my framework as it require specialised knowledge with practical examples.

4.3.2.3. Equity considerations

The equity impact analysis was only performed for the Co-PARS group. The exploratory analysis indicated that Co-PARs was more effective and less expensive for participants: (1) living in the most deprived quintile; and (2) referred for cardiometabolic reasons (e.g. Diabetes, High Blood Pressure and/or Obesity) (Appendix C.4). In terms of age, Co-PARS was more effective but more expensive in older participants (those aged 55 years and over), under NICE's threshold the older cohort were more cost-effective (£2,033 per QALY).

Reflections on item 1.5- What components make up a well-defined study question? (Target population and subgroups)

As recommended in my framework the target population for the economic evaluation were the same participants who were recruited for the effectiveness evaluation. I felt that this was an efficient approach as it meant only one recruitment protocol was required for the trial team. In the initial framework.

Reflections on item 13- What equity subgroups can be considered?

I worked with the PhD student (BB) and trail manager (PW) of the Co-PARs trial to specify in the protocol what baseline data would need to be collected in order to permit these subgroup analyses. The team was not planning to collect baseline data on socioeconomic status, I therefore explained why I felt this was an important piece of data to collect having studied the Marmot Review (Marmot et al., 2010) during my Masters in Public Health in 2012. I suggested to the team that we could collect socioeconomic status data by asking participants to report the first part of their postcode as then I could map the postcode to England's Index of Multiple Derivation (IMD) (Department for Communities and Local Government, 2015). I had come across the IMD while working as a Research Assistant in Public Health. The results of the exploratory subgroup analysis for the Co-PARS sample indicated that future trials might benefit from explore the impact of PA interventions on the those living in the most deprived areas and those who are referred for cardiometabolic conditions.

As specific challenge I came across when carrying out the subgroup analyses, was that due to the trial's small sample size I had to be pragmatic when defining my subgroups. Clearly specified definitions for the subgroups was not something I had considered when drafting the initial framework. Due to the small sample, I decided to categorised the equity subgroups into binary variables. The PA literature was reviewed in order to establish how the equity variables could be categorised into a binary variable. Key UK literature informed the following binary variables: females vs males, aged 55 or over vs under 55; cardiometabolic condition vs other condition, most deprived quintile vs the four least deprived (Scholes, 2017). For socioeconomic status I used deprivation quintiles. The way I categorised this variable into a binary variable was in part influenced by the sample sizes of the groups. For instance, there were a very small number of observations in each of the four least deprived quintiles. I therefore aggregated participants from the four least deprived quintiles to form one equity group and compared this with participants from the most deprived quintile.

4.3.3. Cost consequence analysis

The CCA balance sheet is presented in Table 9 and provides a breakdown of the mean costs and consequences at 6 months (unless stated otherwise) from the difference perspectives: research institute, primary and secondary healthcare, employer, employee. Appendix C.4 provides a more

detailed breakdown of the various costs and consequences. This balance sheet as well as the results in Appendix C.4 can be considered in conjunction with the primary outcome of the effectiveness evaluation, which has been reported elsewhere (Buckley et al., under review).

Table 9. Cost consequence balance sheet

Costs					
Public sector perspective					
Variable	Co-PARS Mean (SE) at 6 months	Usual Care Mean (SE) at 6 months	Control Mean (SE) at 6 months	Incremental difference at 6 months (per participant): Co-PARS vs Usual Care	Incremental difference at 6 months (per participant): Co-PARS vs Control
Leisure Centre operating costs per participant*	£130.53	£76.55	£0	+£53.98	+£130.53
Leisure Centre set up costs (one-off cost, all participants)	£464.10 (All participants)	£0	£0	+£464.10 (All participants)	+£464.10 (All participants)
Research Institute set up costs (one-off upfront costs)	£1,271.16 (All participants)	£0	£0	+£1,271.16 (All participants)	+£1,271.16 (All participants)
Primary healthcare per participant*	£191.65 (£34.52)	£96.17 (£43.01)	£94.66 (£46.53)	+£95.48 (95% CIs:-£40.82 to £231.78; p-value=0.267)	+£96.99 (95% CIs:-£47.37 to £241.35; p-value=0.307)
Secondary healthcare per participant*	£1,326.03 (£387.63)	£1,370.21 (£483.15)	£1,129.88 (£517.71)	-£44.18 (95% CIs:-£1,579.60 to £1,491.25; p-value=1.000)	+£196.15 (95% CIs:-£1,40.65 to £1,801.96; p-value=1.000)

Private perspective					
Variable	Co-PARS Mean (SE) at 6 months	Usual Care Mean (SE) at 6 months	Control Mean (SE) at 6 months	Incremental difference at 6 months (per participant): Co-PARS vs Usual Care	Incremental difference at 6 months (per participant): Co-PARS vs Control
Total employer's loss productivity (numerator/ denominator)#	£36.28# (total for 2 participants out of 15)	£14.51#(total for 1 participant out of 9)	n/a	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Total participants loss of earnings (numerator/ denominator); type	£14.51**(Not mean as only n=1/15)	£0 (n=0/9)	n/a	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Total participant private healthcare (numerator/ denominator); type	£262.40 (n=4/25) Acupuncturist; Chiropractor; Podiatrist	£0 (16/16)	£270.00 (n=1/14) Sports Massage	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Total participant private healthcare (numerator/ denominator); type	£541.20 (n=3/25); Acupuncturist; Counsellor; Podiatrist	£447.99 (2/16); GP consultations; Podiatrist	£655.5 (n=2/14); Physiotherapist; Chiropractor; Sports Massage	Insufficient data for comparative analysis	Insufficient data for comparative analysis

Willingness to pay per induction with ERP for hypothetical PA intervention	£8.22 (£1.52)	£9.25 (£1.90)	£19.64 (£15.58)	-£1.02 (95% CIs:-£7.03 to £4.98; p-value=1.000)	-£11.42 (95% CIs:-£17.70 to -£5.15; p-value=0.000)
Willingness to pay per face to face consultation with ERP for hypothetical PA intervention	-£4.64 (£1.62)	£6.51 (£2.02)	£13.23 (£2.16)	-£1.88 (95% CIs:-£8.27 to £4.52; p-value=1.000)	-£8.60 (95% CIs:-£15.26 to -£1.97; p-value=0.007)
Willingness to pay per telephone consultation with ERP for hypothetical PA intervention	£0.64 (£0.88)	£1.91 (£1.10)	£6.50 (£1.17)	-£1.27 (95% CIs:-£4.80 to £2.21; p-value=1.000)	-£5.86 (95% CIs:-£9.49 to £2.24; p-value=0.001)

Consequences					
Public sector perspective					
Variable	Co-PARS Mean (SE)	Usual Care Mean (SE)	Control Mean (SE)	Incremental difference at 6 months (per participant): Co-PARS vs Usual Care	Incremental difference at 6 months (per participant): Co-PARS vs Control
HRQoL mean 6 Month score*	0.783 (0.029)	0.708 (0.035)	0.767 (0.040)	+0.075 (95% CIs:-0.036 to 0.187; p-value=0.305)*	+0.016 (95% CIs:-0.111 to 0.144; p-value=1.000)*
Change in prescribed any medication	Improvement, 2 less from n=14/16 at baseline	No change from n=12/13 at baseline	No change from n=9/10 at baseline	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Change in High Cholesterol prescriptions	Improvement, 1 less from n=4/16 at baseline	No change from n=4/13 at baseline	No change from n=2/10 at baseline	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Change in Antidepressants prescriptions	Improvement, 1 less from n=3/16 at baseline	No change from n=3/13 at baseline	Worse, from n=1/10 at baseline to 2/10 at 6 Months	Insufficient data for comparative analysis	Insufficient data for comparative analysis
Change in Moderate to strong painkillers prescriptions	Improvement, 1 less from n=1/16 at baseline	No change from n=12/13 at baseline	Worse, from n=1/10 at baseline to 2/10 at 6 Months	Insufficient data for comparative analysis	Insufficient data for comparative analysis

Private perspective					
Variable	Co-PARS Mean (SD)	Usual Care Mean (SD)	Control Mean (SD)	Incremental difference at 6 months (per participant): Co-PARS vs Usual Care	Incremental difference at 6 months (per participant): Co-PARS vs Control
Participant incurred an out of pocket for anything PA-related, excluding induction fee (numerator/ denominator)	68.00% (n=17/25)	80.00% (n=12/15)	57.14% (n=8/14)	Insufficient data for comparative analysis	Insufficient data for comparative analysis

*adjusted for baseline value of that same variable; base on UK's average earnings and hours (unit cost calculation in Appendix C.2). Participant only reported missing 1 hour in work; #Calculations based on consultation frequency and duration of consultations reported.

Reflection on item 1.3- What components make up a well-defined study question? (Secondary Analysis)

I feel the results of the cost-consequence analysis (CCA) (Table 9) make it easier for different stakeholders to identify how the different cost categories are impacted. I became aware of the value of the CCA in January 2019, when I attended a meeting with staff from the local authority. The staff who attended the meeting were involved in the commissioning of the ERS across Liverpool's local authority leisure centres. At the meeting I presented and discussed some of the preliminary results of my Co-PARS analysis. It was through this meeting that I became aware that the staff from the local authority were interested in the results which related to them. More specifically, rather than the total overall cost, they were interested in the breakdown of the results by cost category. In particular, they asked me about how the Co-PARS intervention had impacted on medication use and GP visits during the trial period. The most senior staff member at the meeting informed me that prescription costs was an important topic for local authorities. I also presented a slide about the potential long-term costs which could be made to the NHS if the Co-PARS intervention reduced the number stroke and coronary heart disease events, and incidents of type 2 diabetes. Although, the local authority staff acknowledge this was important, there was less discussion about this slide compared to the cost savings which could be made over the short-term (i.e. during six month trial period). The literature suggests local authority public health teams in the UK are under pressure to contain costs due to the nature of the short-term budget cycles they operate within (Bryan and Williams, 2014). This made me aware of the importance of presenting: (1) the short-term costs and effects; and (2) costs by perspective type rather than a single total cost which value incorporates the costs from all organisations.

4.3.3.1. Intervention costs

From the leisure centre's perspective (in this case, a public sector agency) the operating costs for Co-PARS were £53.98 and £130.53 more per person compared the usual care and control group respectively (Table 9). Total set up costs were calculated for the Co-PARS intervention only. There were estimated to be a total of £464.10 and £1,271.16 from the Leisure Centre and Research Institute's perspectives, respectively. Appendix C.1 provides a more granular breakdown of the intervention costs incurred by the leisure centre (operation and set up costs), research institute (set up costs) and participant (out of pocket costs). This breakdown makes it clear that the largest portion of total costs is attributable to human resource time to deliver the consultations.

Reflections on item 8.1- How can costs be valued? (Intervention costs)

All unit costs for the intervention resources were calculated using published sources. However, not all resource use items have a published unit cost, or a unit cost, which is easy to assign. Each unit cost for the intervention resources therefore required several calculations. These calculations and the assumptions associated with the calculations were recorded in an Excel spreadsheet (Table 2 in Appendix C.2). As a budget breakdown was not available from the local authority, I draw on the Agenda for Change salaries for a range of staff salaries. For example, I used the Agenda for Change Band 4 Fitness Instructor salary as a shadow price for the salary of the Exercise Referral Practitioner at the local authority leisure centre.

Originally, I had considered asking the intervention staff if they were happy to confirm their salaries. Nonetheless, the more I discussed and read in the literature about costing approaches I decided this was not the most appropriate method. More specifically, I discussed costing methods with a group of Health Economists during my study placement at Deakin University. We came to the consensus that it would be inappropriate to ask staff their salaries as some people might feel uncomfortable sharing such information. In addition, an individual's salary may not be representative of the salary observed in standard practice and staff are unlikely to include overhead, superannuation and training costs in their calculations.

Even though I used published unit costs from the NHS (Agenda for Change) it was a time-consuming task which involved a number of analyst-based assumptions. For example, I had to decide whether to assign a cost to the room and IT system that was used for the one-to-one consultations for the Co-PARS and usual care interventions. The consultation room and IT system were existing capital equipment at the leisure centre and so I asked the Leisure Centre Manager to confirm whether the PA intervention meant the room and IT system represented an opportunity cost (i.e. could it have been used for another activity which would have been more beneficial). The Managers at both leisure centre did not believe the use of these resources were displacing other valuable activities and so it was assumed the opportunity cost was small. As a result, I did not assign a unit cost to the room nor IT system. In the initial framework in Chapter 3 I had not considered this. However, this economic concept, known as 'opportunity cost' is an important factor to consider and requires gathering additional contextual information from the intervention providers (hosts).

Overall, I found there was very little published guidance on how to assign costs to the intervention resources without having a budget breakdown available. I also felt the costing approaches I had used in SROI methodology as a Research Assistant in Public Health had not adequately addressed this methodological challenges associated with costing. I

think this is because SROI methodology draws on costing methods from the accounting literature. Some economists feel accounting practices ignore important issues related to costing (Fujiwara, 2015). A key health economic study which I used to support many of my costing decisions was a study from the UK which had been recently published in a high-quality journal, the BMJ (Anokye et al., 2018). I identified this study when I updated my literature search in my systematic review (Chapter 2). The supplementary material for this study provided a detailed breakdown of the calculations performed to calculate the unit costs for the resources to set-up and deliver the intervention, as well as the costs for primary and secondary healthcare utilisation.

4.3.3.2. Healthcare utilisation

At 6 months follow up the Co-PARS participants had incurred around twice as much primary care costs than usual care and the control group. Secondary care costs were greatest in the usual care group (Table 9). Nonetheless, as the wide 95% confidence intervals suggest, there is a lot of uncertainty in these results and the mean differences between the groups were not statistically significant (Table 9). Out of all healthcare activity, the most commonly used service was the GP. Though less common, consultations with a Practice Nurse, Physiotherapist and Counsellor were the next cost drivers (most common and/or greatest unit cost) in primary care. Inpatient admissions had the greatest unit cost across all healthcare activity despite only being accessed by a minority of individuals (see Appendix C.4).

Reflections on item 6.2- How can costs be measured? (Health and social care costs)
Participants were asked to recall their healthcare utilisation over the last 6 months as this aligned with the clinical evaluation data collection time points. Nonetheless, it should be made explicit why healthcare utilisation was not collected at the 12 week data collection point. The one-hour consultation meeting with the team had identified that the questionnaire booklet might be too much of a burden if all questionnaires are asked at all three time points (baseline, and 12-week and 6-month follow up). In addition, studies identified from the systematic review had indicated that it was standard practice to ask participants to recall healthcare use over a 6-month recall period. That said, a consequence of having just two data collection points meant it was not possible to know whether more recent improvements were being overshadowed by high healthcare use in earlier months. In general, completeness of the healthcare utilisation questionnaire was good for most items except for the medication question. For example, a third of the sample failed to report their medication use at follow up. This meant the sample size for complete data was small and so I felt it was inappropriate to present a comparative analysis for the CUA and CCA.

Initial and annual ongoing treatment cost data on CHD, stroke and T2D were expected to come from existing models identified in the systematic review (Campbell et al. 2015; Anokye et al. 2012). When it came to performing the modelling, this was deemed inappropriate for several reasons. These reasons are discussed in the reflection box for item 1.4.

Reflections on item 8.2- How can costs be valued? (Health and social care costs)
Assigning unit costs to the healthcare professionals included in my healthcare utilisation questionnaire involved a number of calculations and assumptions. These were recorded in an Excel spreadsheet for transparency (Table 1 in Appendix C.2). By large I draw on the published unit costs available from the UK's unit cost series for health and social care (Curtis and Burns, 2018). Some published definitions for unit costs in this series differed to the way unit costs were defined in the healthcare utilisation questionnaire and so I had to make some assumptions about what the best shadow price would be.

An additional challenge was that there is no guidance on how to correctly select and assign unit costs for prescribed medications. I referred to the British National Formulary to access specific medication prices, however I observed that assigning a cost involved knowing the dose, quantity and frequency of the medication. Medication data was poorly reported, around a third of participants had not reported medication use despite this information being requested in the questionnaire.

Assigning unit costs for secondary care activity was also challenging. In the UK, there are 2,812 healthcare resource groups (HRG), representing a wide of treatments available in secondary care (NHS Improvement, 2018). It was not possible to know which HRG unit cost was appropriate to assign out of the hundreds of HRGs. This would have required working with secondary care experts to understand what HRG would be most appropriate. My PhD project did not have enough funding to pay a secondary care health professional to work with me to assign HRGs. I therefore addressed this costing issue by using the aggregate average unit costs for outpatient, inpatient, day cases and emergency care visits (NHS Improvement 2018). This is a less accurate approach to costing however it was a quicker and more pragmatic approach, and I ensured I applied the same method to all three comparator groups.

A further issue was that my healthcare utilisation questionnaire only asked participants to name the type of secondary care department they visited and I did ask them to specify whether their hospital visit was an outpatient and day case visits. As a result, I used a mean unit cost (£433) for outpatient and day case visits (NHS Improvement 2018). This was a less accurate but more pragmatic approach.

4.3.3.3. Prescribed medication use

Just over two thirds (n=39/55, 70.91%) of participants reported their medication use across the three groups. The most common prescription type which related to a condition that can be alleviated by PA was for high cholesterol medication. Overall, changes in medication over the 6 month time horizon were small. Table 9, does show a slight trend whereby at 6 months

fewer Co-PARS patients are being prescribed specific medication for a condition which can be alleviated by PA.

4.3.3.4. Participant time and out-of-pocket costs

Loss in productivity was small with just one participant in the Co-PARS group reporting they had a loss in earnings due to taking 1 hour off work to attend the consultation (Table 9). A small number of participants incurred costs for seeking private healthcare across all three groups. Out-of-pocket expenses for PA-related activities and equipment were reported incurred by more than half in each group at both 12 weeks and 6 months. The difference in unit used to report these costs precluded any mean cost calculations (Appendix C.4).

Reflections on item 6.3- How can costs be measured? (Participant costs)

It is unknown whether participants reported the same resource items at 12 weeks and 6 months. In addition, it was not possible to know how often an out of pocket cost was incurred by the participant. My original questionnaire did not ask participants to specify these units (e.g. per event, per month, per 6 months). Around a third of participants did not complete the questions around participant costs meaning it was inappropriate to present a comparative analysis in the CCA. An additional observation was that nearly all participants left the questionnaire's comments boxes blank which suggested that they did not think it was necessary to provide context about the costs they had incurred. Nonetheless, as the analyst, I found it was difficult to interpret this data without any additional units or context.

Reflections on item 8.3- How can costs be valued? (Participant costs)

Prices reported by the participants at 12 weeks and 6 months were used to estimate the costs incurred by the participants. There was a lot of uncertainty in the participant costs which I attribute to the design of the questionnaire I used. In the questionnaire, I had not asked participants to report the frequency of the prices they reported. Therefore, I was unable to tell whether the price reported was a one off payment or a reoccurring cost. This made me recognise the importance of piloting the health economic questionnaires during the feasibility stage of the trial development rather than in the definitive trial.

Reflections on item 6.4- How can costs be measured? (Productivity costs)
Only 25/55 participants correctly answered the question about whether they had lost time in paid work due to participating in the intervention. The chief reason for incorrectly answering the productivity question was due to participants selecting two or more activities (instead of one activity) for explaining how they typically spent their time when they were not taking part in the intervention. Of the 25 participants who correctly completed the intervention, just 20% (n=5) lost time in work in order to take part, with just one participant reporting a loss in earning due to taking part. This suggests that this issue is worth exploring further with a larger sample size to understand how common it is for participants to experience a loss in earnings or give up their leisure time and to understand whether this relates to the outcomes achieved by these participants.

Reflections on item 8.4- How can costs be valued? (Productivity costs)
I had intended to use the UK's average earning from the ONS (2018) to assign a cost for the amount of time participants miss in employment due to taking part in the interventions. Nonetheless, less than half of the participants (45.45%, n=25/55) correctly answered this question and so I reported time lost in work in minutes as opposed to in monetary terms. I felt the sample size was too small and a single cost would be misleading.

4.3.3.5. Change in EQ-5D score

Table 10 reports each groups EQ-5D score for the three time points as well as the mean difference between groups at each time point. This data has also been presented visually in figure 10. The data shows that at baseline the Co-PARS group has a mean EQ-5D score that was lower than the usual care and control group. At baseline this difference was statistically significant (p=0.003) between Co-PARS and the control group, but not statistically significant (p=0.563) between Co-PARS and usual care (Table 10). Figure 10 also shows the three time points, the Co-PARS group had the largest and most sustained improvement in mean EQ-5D score. For usual care and the control group, their mean EQ-5D score fluctuated over time but by a small magnitude.

Reflections on item 7- How can effects be measured?

The EQ-5D-5L questionnaire was straightforward to use. In the Co-PARs study HRQoL was collected at baseline, 12 weeks and 6 months. The area under the curve approach was used to calculate each participants HRQoL utility score over the 6 month period. This score was be combined with life years (the trial follow up period, 0.5 years) to estimate each participants' QALY score. After conducting the trial-based economic evaluation, I had intended to draw on the cohort studies (Hu et al. 2003; 2005; 2007) used in the economic models by Anokye et al. (2012) and Campbell et al. (2015) in order to build a decision-analytical model. Future long-term outcomes were not included due to insufficient data, time and expertise (see reflections in item 1.4)

Reflections on item 9- How can effects be valued?

During the time of the trial, NICE released a position statement where they recommended that the UK's value set for the EQ-5D-5L should not be used to calculate QALYs. NICE recommended that if a study has used this tool (which was the situation I found myself in with the Co-PARS trial) then HRQoL data should be mapped to a published mapping function (NICE 2018). HRQoL scores were valued by mapping EQ-5D-5L scores to the UK's EQ-5D-3L value set through a published mapping function (van Hout, Janssen et al. 2012). This observation highlighted to me the importance of consulting the latest national guidelines for the preferred valuation methods. For the CCA, as recommended all disaggregated effects were reported in their natural units. Reporting the HRQoL scores in their natural units was straightforward.

Table 10. Mean difference in EQ-5D score at 3 time points

Time point for EQ-5D score	Co-PARS Mean (SE)	Usual Care Mean (SE)	Control Mean (SE)	Mean difference (per participant): Co-PARS vs Usual Care	Mean difference (per participant): Co-PARS vs Control
Baseline	0.640 (0.039)	0.724 (0.049)	0.872 (0.053)	-0.084 (95% CIs -0.563 to 0.072; p-value= 0.563)	-0.233 (95% CIs -0.395 to -0.070; p-value= 0.003)
12 weeks score*	0.804 (0.026)	0.754 (0.031)	0.798 (0.035)	+0.051 (95% CIs:-0.050 to 0.151; p-value=0.663)*	+0.006 (95% CIs:-0.107 to 0.120; p-value=1.000)*
6 Months score*	0.783 (0.029)	0.708 (0.035)	0.767 (0.040)	+0.075 (95% CIs:-0.036 to 0.187; p-value=0.305)*	+0.016 (95% CIs:-0.111 to

	0.144; p-value=1.000)*

*adjusted for baseline value of that same variable

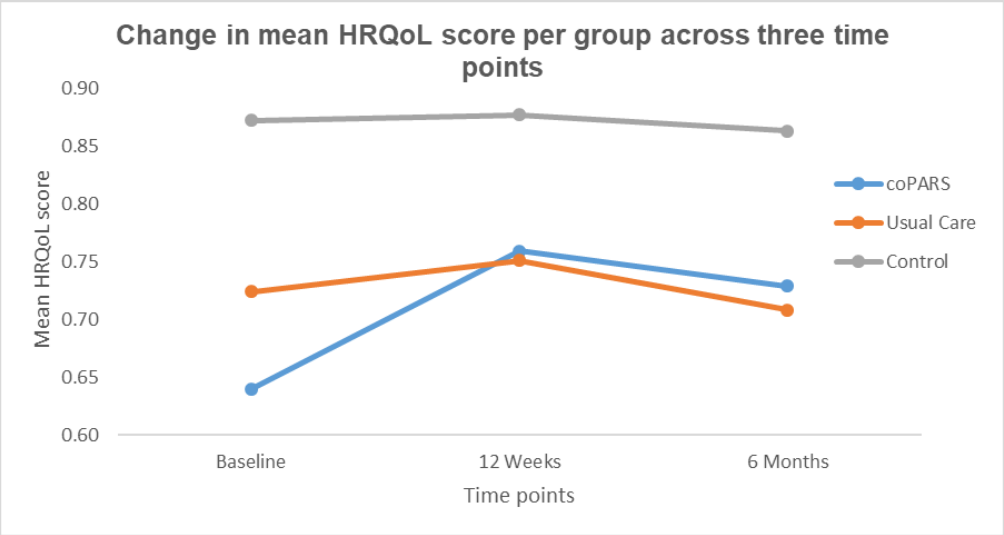


Figure 10. Mean EQ-5D score per comparator group at three time points

4.3.3.6. Willingness-to-pay analysis

At 6 months, participants in the control group were willingness-to-pay more for key components of the hypothetical PA intervention. For example they were willing to pay twice as much as the Co-PARS and usual care groups for an induction and face-to-face consultation. These differences in willing-to-pay was found to be statistically significant when comparing Co-PARS and the control groups prices (Table 9).

Reflections on item 15- How can the results be interpreted?

Methodological choices and any heterogeneity identified are reflected in the strengths and limitations section of the discussion in order to consider how they may have influenced the results and to consider the external validity of the findings. The PRECIS-2 tool was used to support my interpretation of the results (see reflections on item 1.4 for a more detailed explanation of the role of the PRECIS-2 tool in study design and results). In brief, the PRECIS-2 tool goes beyond reflecting on the setting. I found the tool was useful for considering how applicable my results were to standard practice. For instance, I highlight in the discussion section that a strength of this trial was the number of pragmatic featuressuch as: the intervention was delivered by staff who were already employed in the leisure centres, eligibility for the Co-PARs intervention was done by the health professional rather than the researchers, the non-randomised nature reflects standard practice (as participants go to the leisure centre nearest their house) and there was a lack of intensive monitoring of patients to ensure they adhered to the intervention. All of these features are what is likely to happen if the intervention was rolled out standard practice. This means the resource use required if the study was rolled out is likely to be similar to the resource use reported in this study.

Nevertheless, there were features of the trial which did not reflect what would happen in the intervention was rolled out on a larger scale. For instance, the research team coordinated and arranged the set-up meetings. This was initially a hidden resource cost that I had not considered until I came across the inclusion of set-up costs in one of the studies in my review (Chapter 2). The co-ordinating role by the research team was important to support the implementation of the intervention. There is the risk that if these hidden costs are not documented as intervention costs, then the intervention may not successfully be implemented rolled out in the 'real world'. I reflect on this issue of the 'implementation gap' in greater detail in my reflection box for item 3.

Reflections on item 16- How can trial-based economic evaluations be reported?
The Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist was used to write up the methods and results for this study. I found the reporting guidelines particularly useful since I had never conducted nor written up the results of an economic evaluation before. It was therefore a useful checklist for me to refer to as it reassured me that I had reported all the key methods and results of an economic evaluation. Nonetheless, I felt that the explanatory and elaboration notes accompanying the CHEERS checklist would not be appropriate for researchers without training in health economics as the authors use terminology (e.g. cost-effectiveness planes, perspective, discounting) which are not used in the clinical effectiveness literature and so a list of definitions would be required if this checklist is to be used by non-health economic researchers.

4.4. Discussion

The overarching aim of this study was to apply the initial framework from Chapter 3 to the Co-PARS trial. More specifically, the study had two concurrent aims: (1) to reflect on the applicability of the framework to a real world PA trial; and (2) to present an economic evaluations of the Co-PARS trial. The discussion therefore starts with discussion on the reflections on the application of the framework, before going on to discuss the results of the economic evaluation.

4.4.1. Principle findings: Application of the framework

As discussed in the reflection boxes nearly all items in the initial version of the framework require some form of modification. By large, items were not easy to implement in practice as they required further additional calculations and analyst-based judgements to be made, especially in relation to collecting and valuing the resource use data, and interpreting the ICER and uncertainty results. Discussion on how the framework could be improved future application is provided below. These recommendations have been based on the reflection boxes. Reflection boxes in which the recommendations relate to are signposted in parentheses.

4.4.1.1. Primary economic analysis

Additional guidance is required to support analysts conducting economic evaluations alongside trials, notably smaller scale trials and pilot trials where the intervention and/or trial methods are still being refined. The refined framework can recommend that in these circumstances, the primary economic analysis could be to present costs and consequences in a disaggregated format. This should prevent end-users misinterpreting or misusing the point estimates (ICERs) without looking at the results of the uncertainty analysis (see reflections in items 1.2 and 1.3). Similarly, in these circumstances, the refined framework should also explain how modelling may not always be required if there is insufficient robust

data and/or it is anticipated that the benefits of building a model will not outweigh the additional research efforts (costs) (see reflections in items 1.4 and 7.2). Instead, the focus of the economic study should be on generating good quality disaggregated economic data. Systematic reviews can use this disaggregated data to build theory around: (1) how different combinations and levels of resources lead to different levels of outcomes; (2) and how different contexts and patient groups affect this (Anderson, 2010) (see reflections in item 1.4).

A structural modification for the refined framework relates to the recommendation that the summary statistics item (item 11) and uncertainty analysis item (item 14) should be amalgamated into the same item. This is because this study found that the uncertainty analysis was important for interpreting the robustness and generalisability of the results (see reflections in items 1.4, 11 and 14).Overall, the point estimates of the summary statistics presented in this study may be misleading if interpreted alone. As highlighted in the initial version of the framework explanatory notes, if users of the summary decision indices (e.g. ICERs) are not familiar with how these estimates are constructed, then they are less likely to check the results of the uncertainty analyses and demographic data in order to consider the robustness and generalisability of the results, respectively.

Presenting the total mean unit cost for the intervention costs alongside healthcare costs as was done in the CCA, makes it explicit to the decision-maker, how the cost of preventative interventions compare to curative healthcare activity. For example, total cost for the Co-PARS and usual care interventions were £130.53 per person and £76.55 per person, respectively. These costs are negligible compared to the cost of secondary healthcare services, which were the main cost drivers in this study. Average secondary healthcare unit costs for 2018 in the UK were: £3,894 per hospital admission for inpatient care; £742 per day case visit; £160 per A&E visits and £125 per outpatient appointment (NHS Improvement, 2018). To put this in perspective, delivering the Co-PARS interventions to the 25 participants in this study costed £3,263, which costs less than one inpatient admission.

4.4.1.2. Trial design

Economic evaluations are best suited to pragmatic trials (Drummond et al., 2015b). That is to say, the analyst responsible for the economic evaluation should be involved in the trial design from the outset. For instance, during the design stage the health economic analyst should reflect on the nine characteristics of a pragmatic trial, as specified in the PRECIS-2 tool (Loudon et al., 2015) (see reflections in item 3). A key recommendation for the framework is that not only should the analysts consider the PRECIS-2 tool (Loudon et al., 2015) during the design stage (as recommended in the trial design section), but the PRECIS-2 tool can also help in understanding the generalisability of the results after the trial has been conducted (see reflections in item 15). Lastly, as there are ongoing developments in the field of economic evaluation the framework should stress that analysts keep updated on the latest version of their country's guidelines (see reflections in item 9.1)

4.4.1.3. Data collection

The refined framework can suggest that if the trial is constrained by time and resources, the data collection efforts can be prioritised to focus on the most frequently used and/or most costly resource items. Based on this study's findings, the priority resources would be: staff time for delivering and setting up the intervention; GP, practice nurse, physiotherapist and counsellor visits; inpatient, outpatient, day case and emergency visits (see reflections in item 4.3). In particular, for intervention costs, this study found that the TIDieR checklist was a useful tool for carrying out the intervention microcosting exercise (Hoffmann et al., 2014). The framework should also emphasise the importance of understanding how the intervention was delivered in practice from the providers perspective (e.g. the ERPs and leisure centre managers perspectives). This necessitates the researcher to be flexible in their collection approach, in terms of whether this is data is captured prospectively or retrospectively, and via telephone, email or face-to-face (see reflections in item 6.1).

The refined framework can recommend that it is acceptable to ask participants to recall their healthcare utilisation for the previous 6 months (as opposed to more frequently, 12 weeks). In terms of other participant-reported data, since participants' out of pocket costs, time and medication use were reported poorly in this study, there is a need to provide additional guidance around these data collection methods. For participant out of pocket expenses and time costs, the refined framework can recommend that the wording in the questionnaire is improved, so as participants are requested to report the frequency in which they incur a specific out of pocket cost (see reflections in item 8.3). In the present study, comment boxes had been provided so as participants could detail such context alongside their responses, nevertheless these were infrequently used. This means that the framework could also recommend that a subsample of participants could be asked questions about their out of pocket and time costs in interviews/ focus groups to provide clarification, context and an understanding of the acceptability of these out of pocket costs and time costs (see reflections in items 6.3, 6.4 and 8.4). Lastly, as reported in the recent literature, the framework will recommend that it is acceptable to report medication data descriptively, since the data collection methods for medication use is presently underdeveloped, yet public health agencies are interested in data on medication use (Thorn et al., 2018)(see reflections in item 6.2).

4.4.1.4. Assigning unit costs

The refined framework could request analysts to consider whether the capital equipment required for the intervention represents an opportunity cost. This can be done by reflecting on the likelihood that the capital equipment would have been used for an alternative purpose (i.e. is it likely that the private room used for the Co-PARS and usual care inductions would have been used for an alternative purpose, and thus there was a missed opportunity). Questioning whether a resource could have been used for an alternative purpose, which may have generated more benefits, is known as the opportunity cost. This is one of the

fundamental concepts in health economics. If it is unlikely that the capital equipment (private room, IT system, telephone) would have been used for an alternative purpose at the point in time it was used for the Co-PARs and usual care schemes, a unit cost does not need to be assigned to these resource items. Instead, staff overhead costs can be assumed to account for all capital equipment. Nonetheless, all capital equipment should still be reported and quantified in the natural units, so as e.g. leisure centres which do not have private rooms are aware that this resource is required (see reflections in item 8.1). Reporting capital costs in natural units is a strength that will increase the study's applicability to other researchers and decision-makers.

For the present study, matching unit costs to the collected resource use data was a resource intensive task. This was due to the questionnaire requesting participants report their resource use in units of measurement that different to the units in the available published unit costs. The refined framework can therefore recommend that during the questionnaire design stage, the analyst familiarises themselves with the units used in published unit costs. It was also challenging to assign an appropriate cost to the secondary care activity. This is because firstly, participants only reported the departments they visited and secondly in the UK, there are 2,812 unit costs available for the various healthcare resource groups (HRGs). The refined framework can recommend that if analysts are unfamiliar with the HRGs in their country and do not have detail on the exact treatment received by the participants, then a less precise estimate can be used. These can be the average unit costs for inpatient, outpatient, day case and emergency visits. Moreover, the framework can recommend that future healthcare utilisation questionnaire items distinguish between outpatient and day case visits. Overall, judging what were the most appropriate unit costs to use for some resource items was challenging since a number of assumptions are required, especially as organisations provide unit costs that draw on difference accounting methods. This challenge was reported in a recent study, which relied on the assumptions and judgement of the analyst when assigning a HRG unit cost (Anokye et al., 2018). Early research efforts to produce a standard costing methodology for assigning unit costs to self-reported healthcare utilisation have not been successful (Busse et al., 2008).

4.4.1.5. Equity considerations

This study demonstrates wider variations in cost-effectiveness between the subgroups within the Co-PARS sample. Although, the results of the equity impact analyses are limited to the Co-PARS group only and by the small sample sizes, they indicate that all four types of equity groups (age, sex, socioeconomic status and medical condition) which were recommended in the initial version of the framework, should continue to be considered in larger future trials. That said, additional guidance is required for analysts wanting to consider equity in small-scale studies. For example, the framework can recommended that additional literature is sought in order define the equity subgroups into binary variables (see reflections in item 13).

4.4.2. Principle findings: Economic evaluation

The primary economic analysis of this study was to evaluate the cost-effectiveness of Co-PARS using cost-utility analysis. The CUA comparing Co-PARS to usual care generated a cost-effectiveness ratio of £15,349 per QALY, which is lower than NICE's maximum willingness to pay threshold of £20,000-£30,000 per QALY. By itself, this point estimate suggests Co-PARS is cost-effective compared to usual care. Nonetheless, this cost-effectiveness result needs to be interpreted with caution, since the results are not robust to the 95% confidence intervals (-£188,650 to £229,599) and uncertainty analyses. There is uncertainty in where the true population ICER falls on the cost-effectiveness plane and the CEAC indicated that there is only a 26.8% probability that Co-PARS will be cost-effective compared to usual care. The shape of the curve suggests these results are being driven by the between group difference in costs, since cost-effectiveness increased as the threshold increased. The CUA comparing Co-PARS to the control group generated a cost-effectiveness ratio of £157,089 per QALY which is much greater than NICE's threshold. This point estimate needs to be interpreted carefully since the 95% confidence intervals (-£16,035 to £374) and uncertainty analyses indicated that these results are not robust. For example, the CEAC suggested that Co-PARS has no chance of being cost-effective compared to the control group. The shape of the curve suggests this may be due to the between group difference in QALYs rather than costs. This is because, as the threshold increased, cost-effectiveness did not increase.

4.4.3. Comparison with other studies

Four of the ten trials identified in the systematic review of Chapter 2 assessed the cost-effectiveness of PA/ exercise on prescription interventions. Just two of these studies conducted a CUA (Leung et al., 2012, Edwards et al., 2013b), with the study by Edwards et al. (2013) being the only UK-based trial assessing this intervention type. The study had just two comparator groups, the intervention and one control group (who received an information leaflet only). The authors reported an ICER of £12,111 per QALY over a time horizon of 12 months, which they reported as being robust with an 89% probability of the ERS being cost-effective at £30,000. These findings differ substantially to the ICER of £157,089 per QALY reported in the present study when Co-PARS is compared to control group who received no intervention. Likely reasons for this relate to the difference in sample size, Edwards et al. (2013) had a much larger sample size of 798 individuals. Previous modelling studies from the UK looking at the cost-effectiveness of ERS over a lifetime horizon reported that their results were not robust. They found their results to be highly sensitive to small changes in the assumptions made around the cost and effect parameters (Campbell et al., 2015b, Anokye et al., 2011).

In particular, Co-PARS and the control group differed in terms of socio-economic status, which meant heterogeneity may have been directly impacting on the cost-effectiveness result. Future studies should aim to recruit more participants from the most deprived

quintiles, particularly in the no treatment control group. This is an important issue since equity is a key objective of public health interventions in the UK, yet still equity is not routinely or systematically addressed in economic evaluations (Chapter 2). Without considering this equity factor, the conclusions on point estimates risk implicitly saying that it is more cost-effective to do nothing in less deprived groups than to do something in more deprived groups. This would mean decision-makers would be unknowingly choosing to disinvest in the Co-PARS population, who are more likely to be deprived individuals. This opposes the increasing recognition of the need to invest in more inequitable groups (Claxton et al., 2015). Claxton suggests that deprived groups are given additional weighing in cost-effectiveness analyses, where decision-makers agree to pay a greater amount of money per QALY for disadvantaged groups. Future trials could aim to stratify their recruitment strategy in order to recruit a greater proportion of participants from the most deprived quintile so as the comparator groups are better balanced.

This study found the Co-PARS intervention costed £131 per person. This is much lower than the very small intervention costs reported in literature. The systematic review in Chapter 2 found that three of the UK-based models draw on the intervention cost (£186 per person) estimate from a study conducted in 2007 (Isaacs et al., 2007). Reasons for why this estimate may be greater than the intervention cost in this study may be due to the type of activities costed. The EXERT trial which Isaacs et al.'s study assesses, evaluates the costs of a leisure based exercise class intervention. The activities they perform their microcosting study for is based on the cost per exercise class, which is estimated to have a unit cost of £8.02 per PA session attended. The unit cost per 30-minute PA consultation for the Co-PARS intervention was £12.90, which is similar. However, Co-PARS participants could only attend a maximum of four follow up consultations, whereas mean attendance at the exercise classes EXERT trail was reportedly 22.08 sessions. Though both conducted in a leisure centre, the activities involved in the EXERT intervention differed to Co-PARS. This highlights the importance of clearly describing the activities of the interventions being costed (Hoffmann et al., 2014). The ERS intervention assessed by Edwards et al. (2013) is £385 per person. Reasons for why this cost is greater than Co-PARS estimate, also relate to the type of activities costed, for instance, national operating and set up costs are included, as well as local authority and capital costs (e.g. room hire, IT equipment). These costs are appropriate due to the study being conducted at a national level as opposed to two leisure centres. The present study excludes set up costs from the £131 estimate for the Co-PARS intervention, it is deemed appropriate to exclude one-off training from the primary analysis (van Lier et al., 2017) and estimate the intervention as operating in a 'steady state' (Gao et al., 2018, Vos et al., 2007).

In terms of mean EQ-5D scores reported in other studies, the study by Edwards et al. (2013) reported an EQ-5D score of 0.64 in the intervention group at 12 months. Interestingly, 0.64 was the baseline EQ-5D score for the Co-PARS participants. However, at 6 months this

score had increased to 0.78. Previous work has found that physically activity people are associated with having 0.072 higher EQ-5D score than inactive people (Pavey et al., 2011a). Furthermore, some health economists claim that an improvement of 0.08 represents a minimal clinically important difference (MCID) (Luo et al., 2010). However, caution should be taken when interpreting these results, as there is substantial disagreement amongst health economists around whether a MCID is appropriate when interpreting cost-effectiveness results (ISPOR, 2017). Some argue that even a minor improvement in quality of life can be regarded as being cost-effective if someone is willing to pay for that improvement (Glick et al., 2014). The observed improvement in HRQoL in the Co-PARS group therefore is an important and relevant finding which relates to recent debates in the literature around the short-term gains in HRQoL and should therefore be researched further (Campbell et al., 2015b).

4.4.4. Strengths and limitations

The 95% confidence intervals for the incremental analyses, along with the results of the uncertainty analyses indicated that the results in this study were not robust which may be due to a number of reasons. Firstly, it is possible that the study was not sufficiently powered to detect a difference. Secondly, the study was non-randomised which will have increased the risk of there being random variation between the groups which has not been possible to control for through randomisation. Randomisation was deemed unethical in this study, as it was important that participants could select a leisure centre that was most convenient for them. This was a pragmatic feature of the study which reflected standard practice. Since randomisation is not appropriate, future trials should aim to achieve a larger sample size and more leisure centres. Thirdly, the small sample size in the usual care group (n=16) and control group (n=14) meant it would have been inappropriate to adjust for additional baseline variables. This means there may have been heterogeneity driving the cost-effectiveness results which were not controlled for. For example, the baseline HRQoL results demonstrated that there was a statistically significant difference between the Co-PARs and control group. At baseline the control group had a mean baseline EQ-5D score of 0.87, while Co-PARS mean score was 0.64 group. The control groups score aligned more closely to the reported average for the UK general population which in 1999 was recorded as being 0.86 (Kind et al., 1999). Furthermore, the demographics illustrated that there were slightly larger proportion of deprived and economically inactive participants in the Co-PARS group compared to the other two groups, which may have also been driving the results. Finally, a key limitation of the present economic evaluation was that almost a fifth (19.18%, n=13/68) of participants from the original trial were excluded from the analysis, as they were lost to follow up at the 6 month follow up point. Future studies with a larger sample should compare the characteristics of those participants with missing data to those remaining in the study to assess whether there are key participant characteristics, which may explain why participants were loss to follow up. This is important since the existing literature suggests loss to follow up is more likely to be in the least active, thus those who could benefit from the intervention

the most (Iliffe et al., 2014b). A pragmatic feature of the study was that the trial was set in two leisure centres, the Co-PARS intervention was only ran in one centre, which means at present it is uncertain whether the findings such as the observed improvements in mean EQ-5D score, apply to the Co-PARS leisure centre only, or can be applied to other centres across Liverpool.

The Co-PARS intervention made use of the existing leisure centre staff such as the ERP and receptionist, which was a highly pragmatic strategy. In addition, the usual care intervention was a pre-existing ERS meaning it represented what already happened in standard practice. A further strength of this study is that the participants were not recruited to the trial until they had already been referred to an ERS by a health professional. A strength of the analysis for this study was that an intention-to-treat analysis was conducted which is the preferred approach for economic evaluations, as it means the participants were not intensively monitored to ensure they stuck or adhered to the intervention and therefore reflects what would happen in the 'real world' (Hughes et al., 2016).

4.5. Conclusion

The first aim of the study was to reflect on the application of the initial version of the framework to this Co-PARS trial. The framework was not straightforward implement as around two thirds (n=10/16) of the items needed additional methodological judgements to be made by the analyst. In order to fulfil the aim of providing a multidisciplinary framework that ensures a standardised approach to conducting economic evaluations, methodological improvements are required to collect, value and present cost data in a clear and comprehensive way. The cost-consequence balance sheet addressed some of these challenges. The second aim of the study was to present the results of the CUA. The ICER statistic indicated that at 6 months the Co-PARS intervention was cost-effective compared to usual care, but not compared to the control group. Nevertheless, results from the uncertainty analyses indicate that there is uncertainty in whether the true mean ICER value takes a negative or positive value. The improved EQ-5D score in the Co-PARS participants warrants a larger trial with greater power.

Chapter 5: Application of the initial framework to the evaluation of a multi-component intervention aiming to reduce occupational sitting

5.1. Introduction

5.1.1. Background and rationale

Though the literature on SB has only emerged in the last decade, a high quality meta-analysis published in 2012 has shown that SB is associated with elevated risk of 18 types of chronic disease (Wilmot et al., 2012). In particular, the authors of the meta-analysis identified a strong association between SB and diabetes. This is an important finding, since epidemiological and economic data from 184 countries has demonstrated that the global economic burden for diabetes in adults is substantial with the global cost estimated to be $1.31 trillion dollars in 2015 (Bommer et al., 2017). Authors of a more recent meta-analysis focusing on depression argue that reducing SB is not just important for diabetes, but should also be advocated as a strategy for preventing depression (Zhai et al., 2015). In England, self-report data for SB has also illustrated that high levels of SB are associated with poor mental health (Hamer et al., 2014).

Populations most at risk of high levels of SB include office workers. Evidence suggests office workers spend 70-85% of their working day sitting (Healy et al., 2013, Clemes et al., 2014, Morris et al., 2019). There is consensus amongst experts in SB that during a typical working day, workers should aim to accumulate 2-4 hours of standing or light intensity PA per day at work (Buckley et al., 2015). Consequently, interventions aiming to reduce SB in the workplace have become increasingly important. One strategy for which the preliminary evidence has shown positive findings for is the provision of height-adjustable desks in the workplace (Shrestha et al., 2016). These desks enable workers to break up their sitting time by standing up while they continue to work. As demonstrated in Chapter 2's systematic review, the health economic evidence for height-adjustable desk interventions is scant which provides a clear rationale for the present study (Chapter 2).

5.1.2. Aim

It was not possible to draw on the reflections and recommendations from the Co-PARS trial since this study took place during the same period. Therefore, the overarching aim of this study was to apply the initial version of the framework (see Chapter 3) to a SB trial. More specifically, the study had two specific aims:

(1) to reflect on the relevance and applicability of the framework to a real-world SB trial.
(2) to assess the cost-effectiveness of two multi-component workplace interventions, which aim to reduce occupational sitting amongst highly sedentary office workers (contact centre workers).

5.2. Methods for Aim 1: Reflections

Reflections on the planned and actual application of the initial version of the framework are documented throughout the methods and results sections. The reflections aim to provide valuable insight into of the actions I took in order to develop and apply the framework in

practice. Importantly, the reflections aim to provide explanations for why I believed these actions were appropriate. That is to say, the reflections are based on my experience of applying the initial version of the framework to the SLaMM trial. The process I followed for the reflections was similar to the process outlined for the Co-PARs trial. This is because the trials were carried out concurrently and so it was not possible to modify the methods based on the findings of the Co-PARs study. More specific details related to the process can therefore be found in section 4.3. In summary, the reflections describe the actions I took to implement the framework and try to explain why these actions were deemed appropriate. A key action I took which was time-consuming was documenting my costing calculations an Excel spreadsheet so as I had a record of the complexity decisions and assumptions that are involved in assigning unit costs to resource use items (see Appendix D.3). In addition, my reflections refer to the informal conversations and meetings I had with the different researchers and workplace stakeholders involved in my project. In particular, I arranged a one-hour consultation with my supervisory team and key members from the SLaMM trial team. Key members include the PhD student (AM) and MPhil student (DG) working on the trial and the trial manager (LG) who was also one of my supervisors. All three were selected to be involved in the meeting as they had been involved in the design of the SLaMM intervention content and the setting up of the intervention. In addition, they had expertise in physiology, public health and behavioural science. The key objectives of my consultation meeting were to discuss the identification, measurement and valuation procedures. More specifically, this involved: (1) identifying all perspectives (stakeholders) who could experience a change in cost or effects due to the SLaMM trial; (2) discussing the feasibility of incorporating the resource use and EQ-5D-5L questionnaires in the patient questionnaire booklet and the feasibility of capturing intervention costs; (3) explaining the different approaches to valuation I planned to use; and (4) clarifying roles and responsibilities for the data collection process at the key follow up time points. The meeting was a good example of multidisciplinary working.

5.3. Methods for Aim 2: Economic Evaluation

5.3.1. Trial design

This economic evaluation is part of a larger pilot trial (Morris et al. in preparation) which the researcher (Madeleine Cochrane) contributed to all stages of the trial including trial and intervention design, recruitment, set up, data collection and analysis. The pilot randomised controlled trial (RCT) compared two groups: (1) the Sit Less and Move More (SLaMM) group; and (2) the SLaMM Plus (SLaMM+) group (described in section 5.2.3). The intervention start date was the first working day the SLaMM+ agents received their height-adjustable desks (in July 2018). The primary outcome measure for the trial was occupational sitting time. The primary outcome measure for the present economic evaluation is the ICER which will be presented as cost per quality-adjusted life year (QALY).

Reflection on item 1.1- What components make up a well-defined study question? (costs and effects of two or more groups)

Similar to the process for the Co-PARs trial, I was able to explain to the SLaMM trial team that the economic evaluation would have little impact on the trial design that had already been planned for the clinical effectiveness evaluation. There was consensus with the team that the trial provided an important and efficient opportunity to collect additional data for the economic evaluation.

Reflections on item 3. What is *an appropriate study design for a trial-based economic evaluation?*

My initial framework recommended that the economic evaluation is conducted alongside a pragmatic trial. My understanding of what made a trial pragmatic was primarily based on the trial setting and the intention-to-treat principle. Both the SLaMM and SLaMM+ interventions were set in a 'real world' setting as they were delivered in the workplace. I discuss the design of the SLaMM trial in further detail in the reflection box for item 15.

5.3.2. Participants and recruitment

The target population for the trial was adults (≥18 years old) who worked ≥22.5 hours per week as a contact centre agents who will be referred to as agents from here on. In the UK, agents are defined as workers in a contact centre who respond to customer enquiries via the telephone, email or online chat (National Careers Service, 2019). Agents who had a health condition which would prevent them from standing for bouts of 10 minutes or longer, or were pregnant, were not eligible to take part in the trial (see Appendix D.1 for full eligibility screening form). Recruitment posters and emails were disseminated informing agents about the study during May and June of 2018. During the same time period, agents were scheduled to attend a 15-minute group agent briefs (presentations) to learn more about the study. If agents expressed an interest to take part in the trial at the agent briefs, they were requested to provide their personal phone number and/or email so they could be contacted via the research team, to be assessed for their eligibility to take part. All eligible agents were scheduled to complete a 60-minute baseline data collection session during their working hours. At baseline data collection, agents received a participant information sheet and full written consent was obtained. Following baseline data collection, a random number generator was used to randomly allocate trial participants to the two comparator groups (SLaMM and SLaMM+). Ethical approval was obtained from Liverpool John Moores research ethics committee (16/SPS/033).

Reflections on items 1.5 and 13- What components make up a well-defined study question and what equity subgroups can be considered? (Target population and equity subgroups)

The target population was highly sedentary contact centre workers which aligned with the population group targeted for the effectiveness evaluation. The protocol for the effectiveness evaluation included collecting baseline data on age and sex. Nonetheless, there were an insufficient proportion of males and older adults to conduct an equity impact analyses on these subgroups. Due to human error, baseline data on socioeconomic status and pre-existing medical conditions was not included in the baseline questionnaires. I felt socioeconomic status data was particularly important to collect so I made an additional effort to collect this data at the 12-week follow up time point. I requested participants to report the start of their postcode which I could then map to the Index of Multiple Deprivation map (Department for Communities and Local Government, 2015). I had learnt this technique during my experience as a Research Assistant in Public Health. Overall, I feel my belief around the importance of collecting data on socioeconomic status came from my Masters in Public Health training where I learnt about the Marmot Report, a review of health inequality in England (Marmot et al., 2010).

Only around a third (37.5%) of agents provided their postcode data at the 12-week follow up period, this highlighted the importance of requesting this data at baseline alongside the other demographic variables. Due to the limited amount of data on socioeconomic status, it was not possible to do a subgroup analysis according to area of deprivation. As an alternative, I conducted a post-hoc exploratory analysis for socioeconomic status where I used education level as an indicator for socioeconomic status. I made the assumption education was an appropriate indicator as evidence from the UK shows that those with university degrees have better health and live longer than those without (Marmot, 2010).

5.3.3. Comparator groups

5.3.3.1. Sit Less and Move More (SLaMM) intervention

The SLaMM intervention was a 12-week intervention comprising of the following key components, which included: three 30-minute education and training sessions about the benefits of, and ways to, reduce SB in the workplace; 12 weekly infographic emails promoting increased movement in the workplace; and a timer and daily goal-setting log book. Examples of the weekly emails and daily logbook which the researcher (Madeleine Cochrane) contributed to are provided in Appendix D.1.

5.3.3.2. Sit Less and Move More Plus (SLaMM+)

The SLaMM+ intervention was a 12-week intervention comprising of the same key components as the SLaMM intervention, but with the addition of a height-adjustable desk. All SLaMM+ agents were provided with their own personal height-adjustable desk. These desks

are intended to allow participants to carry out their work in a seated or standing position with the flexibility to alternate between the two options during the working day.

5.3.4. Type of economic evaluation

This economic evaluation compares the costs and consequences for the two intervention groups over a 12-week time horizon. The primary analysis of this study was to conduct a trial-based CUA from a multi-agency UK public sector perspective. The secondary analysis was to present a CCA. A key methodological feature of this study is the piloting of the initial framework. Table 18 (section 5.3.4) illustrates how the framework was intended to be applied to the present study along with the reflections on how it was actually applied.

5.3.4.1. Perspective

As the trial was set in the workplace, a CUA was conducted from a multi-agency public sector perspective, as recommended in the UK's guidelines for the assessment of interventions delivered in non-healthcare settings (NICE, 2014a). The multi-agency public sector perspective included the research institute (payer), and primary and secondary healthcare agencies. The CCA included the same public sector agencies listed for the CUA, as well as the perspective of the private agencies/ agents which included the employer and employee (participant). Cost categories deemed relevant for each perspective and economic evaluation type are outlined in Table 12. Rationale for the choice in cost categories is provided in Chapter 3.

Table 11. Perspective, cost categories and economic analysis

Sector	Perspective	Cost category	Economic evaluation type
Public	Research Institute (Payer)	Intervention operating costs	CUA; CCA
		Intervention set up costs	CCA
	Healthcare sector	Primary healthcare	CUA; CCA
		Secondary healthcare	CUA; CCA
		Prescribed medications	CCA
Private	Employer (Host)	Intervention operating costs	CCA
		Intervention set up costs	CCA
		Absenteeism	CCA
		Presenteeism	CCA
	Agents (Participants)	Time costs	CCA
		Travel costs	CCA
		Out-of-pocket costs	CCA

CUA – cost utility analysis, CCA – cost consequence analysis

5.3.4.2. Data collection procedure

The trial had two key data collection points. Baseline data collection took place 2-4 weeks prior to the start of the intervention. The follow-up data collection point took place 12 weeks

after the start date. All data collection took place at the contact centre during the agents working hours. Demographic measures reported in this study were collected at baseline. Economic measures were collected at both baseline and 12 weeks. The primary economic researcher (Madeleine Cochrane) supported the wider collection of data and in particular, was responsible for coordinating and handling the collection and analysis of all economic data. This included the data from the intervention microcosting exercise, which were collected during the trial period. Additional outcome measures such as occupational sitting time and a range of physiological measures were collected as part of the clinical effectiveness trail, but are reported elsewhere (Morris et al., in preparation).

5.3.6. Cost measures

Two types of data were required: (1) quantities; (2) unit costs. For the present study, primary data collection methods were used to estimate resource use quantities, while unit costs came from secondary sources. Measurement tools used to capture resource use quantities are outlined below and described in Appendix B.1. Research costs were not included in the study. Costs categories included are detailed in the subsequent sections.

Reflections on item 6.2- How can costs be measured? (Health and social care costs)
At baseline and 12-weeks follow up, the self-report healthcare utilisation tool presented in Appendix B.1. was be used. Participants were asked to recall their healthcare utilisation over the last 12 weeks to align with the clinical evaluation data collection time points. Just under a fifth (17.5%; n=7/40) did not complete the medication question. Participants left this question blank which made it difficult for me know whether it was blank because it did not apply to them or because they could not recall the medication they were being prescribed. As discussed in the reflection box for item 1.4, I did not have the time or expertise to build a model to estimate the long-term effects of the SLaMM+ and SLaMM intervention. Future healthcare treatment costs were not therefore included.

Reflections on item 6.3- How can costs be measured? (Participant costs)
Participant out of pocket costs were requested in the 12-week follow up questionnaire (see Appendix B.1). In future studies in similar workplaces, if questions need to be removed from the questionnaire booklet in order to reduced participant burden then this would be a question that could potentially be removed. I recommend this because no participants from the SLaMM trial reported experiencing out of pocket costs due to taking part in the intervention. I expect this is likely to be due to the nature of the SB trial which involved encouraging low intensity PA in the workplace (e.g. increasing standing and walking during working hours).

Reflections on item 6.4- How can costs be measured? (Productivity costs)
I obtained objective absenteeism and presenteeism data (company-specific job performance metrics) from the employer for the 12 weeks prior and 12 weeks after baseline. The objective job performance data (presenteeism) provided by the employer was incomplete and was difficult to make a meaningful comparison from. Therefore this data was not presented in the CCA. The data was complex to interpret because different agents had been assigned different performance targets depending upon which specific job contract (job role) they were employed on that week. In total participants were working across three types of job contracts. In addition, some agents changed contracts multiple times over the 24 week period (12 week before and after baseline) making it difficult to identify a trend and change in their productivity levels. I was surprised by this finding since staff at the company had mentioned on several occasions about the importance of the productivity metrics. I discussed this with the Centre Contracts and they explained that because there was a high turnover of staff at the company, with around a third of agents leaving within the first three months, for some agents the team managers monitored the individual's immediate day-to-day metrics, rather than the longer term metrics (e.g. 12 weeks). Overall, understanding the employer's productivity data, linking up multiple datasets which comprised of the productivity data and deciding how informative the data would be was a time-consuming exercise. This experience made me consider if this was the reason why I did not identify any studies through my systematic review (Chapter 2) which had included objective measures for presenteeism. By contrast, the self-reported presenteeism questionnaire by Lerner et al. (2001) had a 100% (n=40/40) completion rate and provided a standardised way to measure presenteeism. In terms of objective company absenteeism data, this data was incomplete for 25% (n=10/40) of agents despite my additional efforts being to retrieve this data from the company. Overall, this experience made me recognise the complexity and challenge associated with accessing robust cost data from data sources which are not set up for research studies.

5.3.6.1. Intervention costs: quantities

The trial protocol and CONSORT flow diagram were firstly drawn on to identify key intervention activities for each comparator group (Appendix D.2). In addition, more precise prospective microcosting methods were employed during the 12-week intervention period. The researcher (Madeleine Cochrane) and the Resource Planner from the Contact Centre recorded the type and price of the equipment, and amount of time that was spent on delivering and setting up the intervention using weekly electronic logs. Structure of the weekly electronic log was informed by the diaries used a previous public health microcosting study (Charles et al., 2013). Retrospective microcosting methods were also performed, this included a one-hour consultation with two centre contacts at the company where the intervention was hosted. The centre contacts were key to the set up and delivery of the intervention. More specifically, they were two members of staff employed by the workplace

who were selected by the workplace's senior management team to be the point of contact for the researchers. The centre contacts liaised with the researchers in order to support the implementation of the interventions. Therefore, they were requested to recall and estimate the amount of time and/or equipment they had invested in delivering and setting up the 12-week intervention. The consultation drew on the template for the weekly electronic log as well as the items for intervention description in TIDieR framework (Hoffmann et al., 2014). All data was recorded in a single microcosting spreadsheet (see Appendix D.2 for an example on how the microcosting tool was populated).

Reflection on item 4- What costs are important and relevant?

In item 4 of the initial framework I provide examples of five perspectives that may be relevant to the analysis of a SB intervention: the payer, the provider, health care, the participant and the employer. As I was involved in the implementation and delivery of the SLaMM trial, I was able to identify who the payer and provider was through my own observations. That said, I had expected the employer organisation (call centre) to predominantly represent the payer, provider and employer perspective. To my surprise the costs associated with the employer organisation in the SLaMM trial was much more complex. I documented this complexity through my daily observations. I feel the CCA provides a summary of this complexity as it shows which stakeholder incurs which costs. For example, the breakdown in the resource costs in the CCA reveals that the average SLaMM+ operating costs incurred by the employer were similar to the average SLaMM+ operating costs incurred by the Research Institute (excluding trial-related research costs). The Research Institute paid for a large proportion of the operating costs of the intervention since the PhD students delivered three the education and training sessions over the 12-week period (week 1, 3 and 10). This is an interesting observation to reflect on in terms of considering who would deliver the education and training components if the intervention was rolled out across multiple organisations. The people delivering the education and training component would require expertise in SB in order to provide this component. Training key staff members from the workplace to be able to deliver this component of the intervention could be one way to achieve this and would incur a one-off training cost.

Another interesting observation through the CCA in relation to relevant costs, is that the average SLaMM+ set up costs incurred by the employer are greater than the average SLaMM+ set up costs incurred by the Research Institute. I had expected that the Research Institute would have a greater role in the set-up that the employer organisation. Through informal data collection methods (face-to-face one-hour discussion and weekly logs) I became aware of the amount of the time Centre Contacts and Resource Planner have input into the set up stage of the SLaMM+ intervention, these included additional meetings and planning time: Installation of height-adjustable desks, a one-hour consultation with senior management at the centre contact, 30 minutes consultation with Resource Planning team, organisation of Team Manager briefs, organisation of call agent briefs (by scheduling offline time for the agents). The Resource Planner explained that scheduling offline time for the agents (study participants) was a time-consuming and complex process, as it required them to schedule enough agents to work ‘offline’ (i.e.not available to answer phonecalls to customers) at the same time so as the meetings did not need to be repeated. Nonetheless, the staff reported that they needed enough agents to be working ‘online’ (i.e. answering phonecalls from customers) in order to avoid the SLaMM trial impacting on the company’s business metrics. This highlighted to me the relevance of capturing the employer’s perspective in economic evaluations assessing

workplace interventions. It also helped me realise that the problems the company were facing related to resource use (in terms of time). It made me realise the it could be helpful to incorporate a greater economic perspective into the design of interventions in order to ensure they are pragmatic and acceptable from a resource (and employer's) perspective.

I had expected loss in earnings to be incurred by the employer only however through the analysis of the demographic and absenteeism data I became aware that the participants also experienced a loss in earnings if they were off work sick due to ill health. More specifically, through the demographic data I learnt that the majority of agents (85%, n=34/40) who took part in the SLaMM trial were agency staff. The Centre Contracts explained that agency staff do not receive pay when they take uncertified sickness. The absenteeism data I analysed revealed that participants were more likely to take uncertified sickness than certified sickness (Appendix D.4). Productivity loss (in terms of loss of earnings) was therefore was an important cost category from the participant's perspective. This highlighted to me the importance in understanding how a company is organised.

Analysis of the healthcare utilisation data revealed which healthcare activities were the cost drivers and could therefore be prioritised in future studies. I refer to cost drivers as the healthcare activities which were more frequently used by participants and/or had a greater unit cost relative to the other activities. For example, GP, Practice Nurse and Counsellor visits were the most commonly used healthcare professionals in primary care. In addition, although secondary care healthcare use was reported less frequently than primary care activities, secondary care activity typically had a much greater unit cost than primary care activity.

5.3.6.2. Intervention costs: unit costs

Published standardised unit costs were used as recommended in the framework. This included published unit costs for all staff's time, travel and printing costs. Salary oncosts (national insurance and superannuation at 14%) were included for all staff except for the agents since these were by large non-permanent staff. Overhead costs for the interventions were minimal, it was therefore appropriate to exclude these costs (Edwards et al. 2019; Drummond et al. 2015). Consequently, all capital equipment (e.g. the private room, IT system, telephone) are not included in the total intervention set up and operating costs. Nevertheless, they were still quantified in their natural units and are reported in the CCA. All unit cost calculations and secondary unit cost sources are detailed in Appendix D.3.

5.3.6.3. Healthcare costs: quantities

At baseline and 12 weeks, participants self-reported their healthcare use over the previous 12 weeks using an adapted version of the widely used healthcare utilisation questionnaire

called the client service receipt inventory (CSRI) (Mayer and Beecham, 2005, Beecham and Knapp, 2001).

5.3.6.4. Healthcare costs: unit costs

As recommended in Chapter 3, national published unit costs were sourced. More specifically, primary healthcare unit costs came from the UK's annual Health and Social Care unit cost publication (Curtis and Burns, 2018) and secondary healthcare costs came from the UK's reference costs database (NHS Improvement, 2018). Calculations and unit cost sources are provided in Appendix D.3. Medication costs were not assigned a unit cost due constraints in the patient-level data, and the researcher's time and expertise. That is to say, prescription medication costs are not included in the total healthcare utilisation costs. Nonetheless, medication quantities are reported in their natural units in the CCA.

5.3.6.5. Employer costs: quantities

Participant's productivity in terms of presenteeism was captured via self-report at both baseline and 12 weeks using the Workplace Limitations Questionnaire (WLQ). Absenteeism from the employer's perspective was quantified objectively using certified sickness data provided by the employer. The company provided data on absenteeism for the 12 weeks before and during the intervention.

5.3.6.6. Employer costs: unit costs

Certified absenteeism was interpreted as being part of the numerator (cost) part of the analysis. By contrast, as the methods for incorporating presenteeism into economic evaluations are underdeveloped, presenteeism was reported in the CCA in its natural units.

5.3.6.7. Participant costs: quantities

An adapted version of the annotated patient costs questionnaire (Thompson and Wordsworth, 2001) was completed by the participants at baseline and 12 weeks (see Chapter 3). This questionnaire asked participants whether they had incurred any out-of-pocket costs due to participating in the intervention. The CSRI healthcare questionnaire (section 5.2.6.4) asked participants to report if they paid privately for any healthcare they accessed. These costs were reported in the CCA but excluded from the CUA, as the CUA was from a public sector only perspective. Absenteeism from the participant's perspective was quantities objectively using uncertified sickness data provided by the employer. The company provided data on uncertified absenteeism for the 12 weeks before and during the intervention.

5.3.6.8. Participant costs: unit costs

Time was reported in natural units (hours/ minutes). Out-of-pocket costs for equipment and leisure centre memberships were reported and valued using the actual prices reported by the participants. Private healthcare utilisation was reported using unit costs from the UK's annual Health and Social Care unit cost publication (Curtis and Burns, 2018). Uncertified sickness, which for most participants was taken for less than 4 days in a row, was

interpreted as being a loss in earnings from the employees perspective. The company provided this contextual information regarding the nature of the agent's non-permanent contracts. Consequently, uncertified absenteeism was costed using the human capital approach, whereby time lost in work (in minutes) due to uncertified sickness was a loss in earnings for the participant.

Reflections on item 8.3- How can costs be valued? (Participant costs)
Participant's costs also included the loss in earnings due to uncertified sickness. A human capital approach was used whereby the mean time missed due to uncertified sickness was matched to average wages. No participants reported any out of pocket costs. Productivity loss was identified as an important participant cost as the majority (85%, n=34/40) of participants were agency staff. The Centre Contacts and Resource Planner explained that non-permanent staff did not receive earnings when they were off sick, unless this was certified sickness or the agent had been sick for 4 or more days in a role. Human capital approach was appropriate as it is a commonly used method which takes the perspective of the sick employee (van den Hout, 2010). Only a minority of agents had uncertified sickness for either: (1) 4 or more working days in a row; or (2) at least 2 working days in a row, which in addition took place the day before or after a non-working day (the weekend). These definitions were based upon the UK's eligibility items for statutory sick pay. There was uncertainty in whether these agents claim statutory sick pay. That is to say, I employed a standardised approach was used, whereby I assumed that all participants with uncertified sickness experienced a loss in earnings for their uncertified sickness hours. This assumption may mean loss in earnings from the participant's perspective are overestimated. For certified sickness for the agency-based staff I used the UK's weekly Statutory Sick Pay (SSP) for 2018/19 (HM Treasury, 2019) to represent the cost provided by the government when a participant took certified sick leave for specific time period. Overall, the costing approach I have outlined here demonstrates the many calculations and assumptions required by the analyst.

5.3.6.9. Currency, price year and conversion

Dates of all prices are reported in Appendix D.3. Nearly all unit costs came from secondary sources for the current price year (2018/19). Adjustments were made to all unit costs before they were multiplied by the resource quantity data. Where the price year diffed, a price-year adjusted cost estimate was calculated. This was done by adjusting unit costs to the target year (2018/19) and by applying the UK's GDP deflator index (HM Treasury, 2019). All unit costs came from UK sources so currency conversions were not necessary. As this is a trial-based economic evaluation conducted over a 12-week time horizon, no discounting was required.

Reflection on item 1.4- What components make up a well-defined study question? (Time Horizon)

In item 1.4 of the framework presented in Chapter 3, I advise that if there are sufficient data, time and expertise, then a decision model can be conducted after the trial-based economic evaluation. There was one main barrier which precluded a decision-model being built for my analysis of the SLaMM. This related to the time it would take to build a decision model from scratch. Through discussions with other health economists during my study placement with Deakin Health Economics group we concluded that building a SB decision model from scratch for the SLaMM economic evaluation would be a time-consuming process and would be beyond the scope of my PhD project. Therefore, rather than building a model from scratch we concluded it would be sensible to discuss the potential long-term impact in a descriptive way.

Reflections on item 10- How can costs and effects be discounted to a present day value?

As I did not include future costs and effects in the analysis I was not required to apply a discount rate. The trial-based economic evaluation assessed costs and effects over a 12-week time horizon only.

5.3.7. Economic outcome measures

5.3.7.1. Quality-adjusted life years (QALYs)

The primary outcome measure for the CUA was quality adjusted life years (QALYs) as recommended in the UK's reference case (NICE, 2014a). EuroQol's validated and widely used generic measurement tool called the EQ-5D-5L (EuroQol 5 dimension, 5 level) was used to measure HRQoL (Rabin et al., 2011). Pre-existing preference weights (HRQoL index scores) for the UK population were matched up to each health state to calculate each participant's HRQoL utility score. This is the preferred method for studies from the public sector perspective in the UK (NICE, 2014a). The final step to deriving the HRQoL scores involved mapping the EQ-5D-5L index scores to the EQ-5D-3L using a recommended mapping function (van Hout et al., 2012). Mapping is recommended in NICE's recent position statement for the valuation of the EQ-5D-5L measurement tool (NICE, 2018). The EQ-5D-5L was included in the questionnaire booklet at baseline and 12 weeks. In order to calculate QALYs for each participant, an average of each participants two HRQoL utility scores was calculated and then combined (through multiplication) with length of life which at 12 weeks was (0.23 years). A breakdown of HRQoL over time was reported in the CCA table.

Reflections on items 5, 7 and 9: What effects are important and relevant and how can they be measured and valued?

As recommended in item 5 of the framework, I used a single generic measure of health benefit for the primary outcome (the QALY). Overall, items 5 and 7 were straightforward to apply. Nonetheless, calculating the length of life part of the calculation required some additional consideration as I needed to calculate what the equivalent of 12 weeks was using the 1 life year metric. In terms of calculating the quality of life side of the QALY calculation, I feel this was straightforward as there is clear consensus on what the preferred methodological approach is for studies conducted from the UK. More specifically, the EQ-5D measurement tool is recommended to measure the HRQoL part of the QALY calculation (NICE, 2014a). Valuing the EQ-5D utility scores was straightforward as I followed the UK-specific guidelines. The reflections I reported in the reflection box for item 9 in Chapter 4 (the Co-PARs study) apply to my experience in the SLaMM trial as both trials were carried out in tandem.

5.3.8. Equity considerations

Demographic equity-relevant data was collected at baseline through the self-report questionnaire booklet. The main demographic information collected aligned with the equity characteristics recommended in Chapter 3: socio-economic status (postcode), age and sex. As there was insufficient postcode data available, tertiary education was used as an indicator for socio-economic status.

5.3.9. Analysis

5.3.9.1. Complete case analysis

The CUA was a complete case analysis and followed the per-protocol principle, whereby participants were excluded from the analysis is they did not adhere to the intervention. For multi-item measurement tools where only a small proportion of the data was missing (less than 10%) it was deemed acceptable to impute the mean of each group for participants missing an item (Eekhout et al., 2014). In the CCA, medication, absenteeism and presenteeism data were reported based on the number of available-cases. Heterogeneity between the groups was assessed through descriptive statistics by comparing the groups' baseline characteristics.

5.3.9.2. Summary statistics

A participant-level analysis was performed, where costs and QALYs for each participant were presented. Total mean costs were calculated using the absolute intervention and healthcare costs incurred between baseline and the 12 week follow up period. Area under the curve between baseline and 12 weeks was used to estimate the change in QALYs. SLaMM+ mean costs and QALYs were compared to the mean values for the SLaMM group. The incremental cost-effectiveness ratio (ICER) was estimated. Measures of sampling variability are presented alongside the point estimates (standard deviations and 95%

confidence intervals). Precision in the mean estimates was improved by adjusting for baseline differences in the dependent variable using multiple regression. Results based on unadjusted estimates are presented in Appendix D.4.

5.2.9.3. Stochastic uncertainty analysis

Stochastic uncertainty was assessed through a bootstrapping simulation of 1,000 bootstrapped replicates. From here, cost-effectiveness planes and cost-effectiveness acceptability curves were produced to help show the uncertainty in the summary statistics. In addition, a one-way scenario analysis was performed to consider the variation and uncertainty in the total cost estimate when the employer's intervention and certified absenteeism costs were included in the CUA.

5.4. Results

5.4.1. Baseline characteristics

Figure 11 shows the number of agents invited, eligible and enrolled to take part in the trial between May-July 2018. In total, 60 agents were enrolled onto the trial, half (n=30) were randomised allocated (through a computer randomisation system) to the SLaMM+ intervention and provided with a height-adjustable desk. Forty participants completed the 12-week follow up assessment and were included in the complete case analysis. As highlighted in Figure 11, the most common reason (70.58%, n=12/17) for discontinuing with trial, was due to the agent leaving the company.

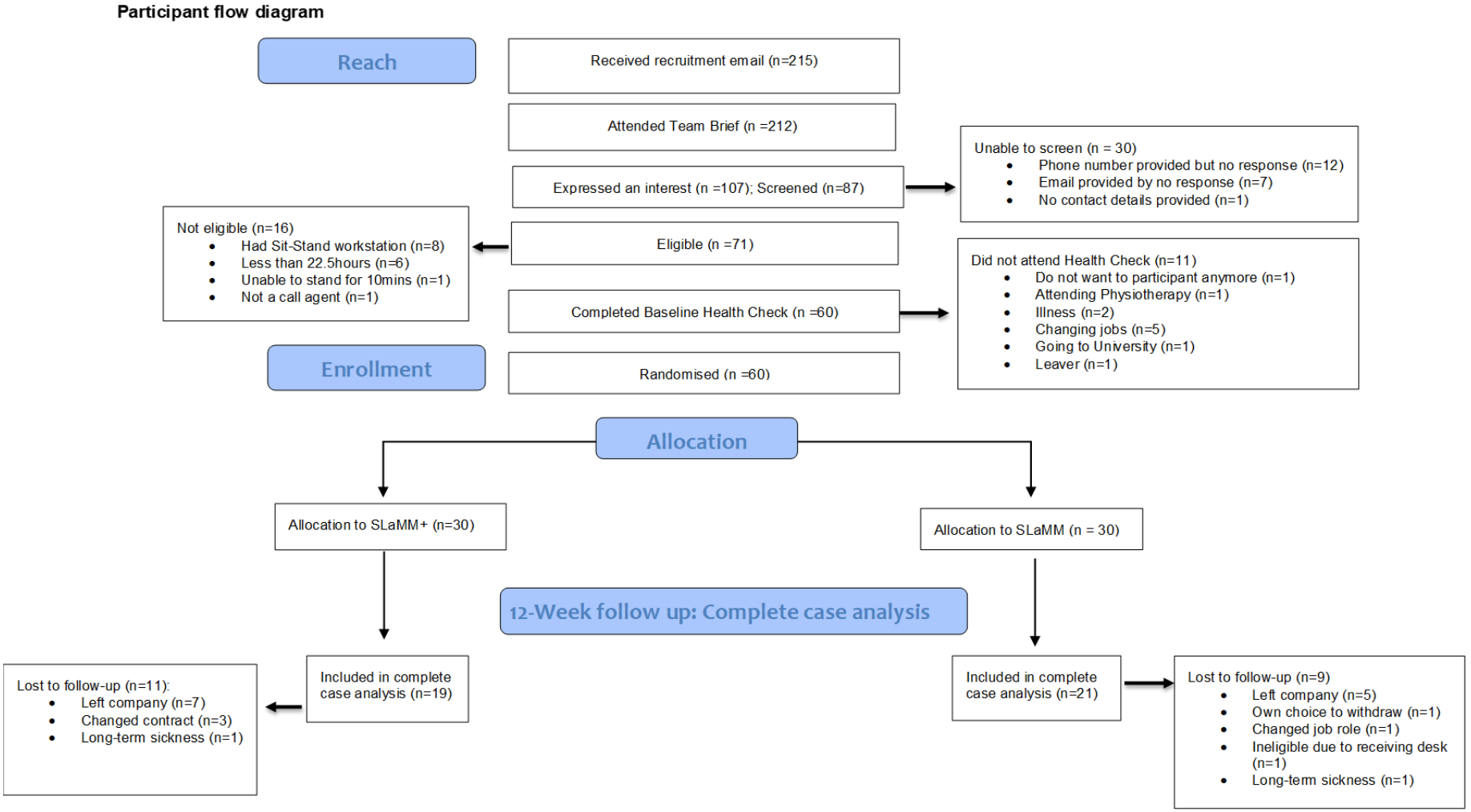


Figure 11. Flow diagram for participants enrolled on the trial

Table 12 details participant characteristics for the economic sample. In general, the characteristics of the two intervention groups were similar. As noted in the table, the SLaMM+ group had a slightly greater proportion of white British participants, males, smokers, non-binge drinkers and agency contracted employees. As socioeconomic status was excluded from the baseline and 12-week follow up questionnaire, this data was requested retrospectively. Just over a third (37.5%, n=15/40) provided this information.

Table 12. Baseline characteristics

Variable	SLaMM+ (n=19)	SLaMM (n=21)
Tertiary educated	52.9% (n=9/17)	47.6% (n=10/21)
Ethnicity: White British	100% (n=18/18)	85.7% (n=18/21)
Smoker	55.6% (n=10/18)	28.6% (n=6/21)
Average number of fruit and vegetables eaten per day	2.4(±0.55) (n=15/18)	1.8(±1.49) (n=18/21)
Binge drink	0.0% (n=0/12)	16.7% (n=3/18)
Employment status: non-permanent agency contract	100% (n=18/18)	76.2% (n=16/21)
Sex: Female	68.4% (n=13/19)	81.0% (n=17/21)
≤1 year at the company	88.9% (n=16/18)	81.0% (n=17/21)
Mean age (years)	27.7(±9.5) (n=18/18)	31.3(±11.1) (n=21/21)
Aged 55 or over	0.0% (n=0/18)	4.8% (n=1/21)
Work full-time (37.5 hours per week)	88.9% (n=16/18)	81.0% (n=17/21)
Live in 20% most deprived area nationally	100% (n=7/7)	75% (n=6/8)

5.4.2. Cost-utility analysis

5.4.2.1. Summary statistics

Cost and QALYs based on the adjusted analysis are presented in Table 14. At 12 weeks mean incremental QALYs were higher in the SLaMM+ group (+0.006; 95% CI:0.190 to 0.206) compared to the SLaMM group. Mean incremental costs were also higher in the SLaMM+ group (+£228.55; £-978.65 to £521.55). In summary, the SLaMM+ group costed more, but also experienced more QALYs. In terms of cost-effectiveness, the ICER was £38,091 per QALY, which under NICE's WTP thresholds suggests SLaMM+ is not cost-effective compared to SLaMM. In all incremental analyses, the 95% confidence intervals crossed zero indicating the results were not statistically significant (Table 14). The results of the CUA based on all unadjusted patient-level data are presented in Appendix D.4.

Table 13. Results of CUA

Variable	SLaMM+ Mean at 12 weeks (SD) per participant	SLaMM Mean at 12 weeks (SD) per participant
QALYs*	0.198 (SE:0.004; 95% CI: 0.190 to 0.206)	0.192 (SE: 0.004; 95% CI: 0.185 to 0.199)
Total costs*	£518.65 (SE: £265.78; 95% CI: -£19.88 to £1,057.17)	£290.10 (SE:£252.56; 95% CI: -£221.64 to £1,057.17)
Incremental QALYs & Costs *		
Incremental QALYs: SLaMM+ vs SLaMM	0.006 (SE:0.005; 95% CI: -0.005 to 0.017; p-value=0.266)	
Incremental Costs: SLaMM+ vs SLaMM	£228.55 (SE:£370.20; 95% CI: £-978.65 to £521.55; p-value=0.541)	
ICER statistic at 12-weeks*		
ICER point estimate	£38,091 per QALY	
ICER 95% CIs at 12 weeks**: SLaMM+ vs SLaMM		
95% CI for ICER based on 1,000 bootstrapped simulations	-£254,156 to £173,247	

*Adjusted for baseline imbalances in HRQoL score and healthcare costs; ** original unadjusted data

*Reflections for item*s *1.2* and *11- What components make up a well-defined study question (primary analysis)? And what summary statistics can be presented?*
The primary analysis of this study was to conduct a CUA. This aligned with the recommendations I made in my framework in Chapter 3. My reflections on whether the CUA was most appropriate for the primary analysis are similar to the reflections I have made for the Co-PARs trial (see the reflection box for items 1.2 and 11 in Chapter 4). As the Co-PARS and SLaMM trial ran concurrently my reflections on the Co-PARS analysis and results had not been systematically studied in time in order to inform the SLaMM analysis.

Overall, I felt the ICER result by itself did not inform the team about the uncertainty in the data that underpins the ICER calculation. When I reported to the team who had been involved in designing the SLaMM trial that the ICER was above £20,000 per QALY (Table 14) I found it difficult to explain the uncertainty associated with the ICER statistic, which was represented in the cost-effectiveness planes, and cost-effectiveness acceptability curves I produced. I showed the uncertainty analysis to my three supervisors who were non-health economic researchers. I was aware that these concepts were difficult to understand without training in health economics as they are not used in the clinical effectiveness literature. Prior to completing the Applied Methods of Cost-effectiveness Analysis course at the University of Oxford in 2018, I too had felt that I did not completely understand these concepts. It was only through doing practical exercises that I felt more confident in the meaning of the uncertainty results. This made me realise it would be difficult for my framework to include guidance to non-health economic researchers on how to carry out an uncertainty analysis for an economic evaluation.

Nevertheless, I was able to explain to the team that the small sample size may have lead to the uncertainty observed in the summary statistics for the mean cost and QALY data, as shown through the 95% confidence intervals (Table 14). 95% confidence intervals are commonly discussed in the effectiveness literature and so this was easily understood by the team. The team understood that the large confidence intervals may be related to the small sample size and therefore the results need to be interpreted cautiously. It is not possible to calculate confidence intervals for ratio statistics such as the most commonly used cost-effectiveness summary statistics, the ICER statistic.

Reflections on item 12- What adjusted analyses can be performed?
The12-week costs and QALYs were adjusted for baseline imbalances in healthcare costs and HRQoL scores using multiple regression. As discussed in Chapter 4 for the Co-PARs study, I felt confident in the adjusted analysis I performed (Table 14) as multiple regression is a widely used statistical approach within the effectiveness literature.

5.4.2.2. Uncertainty analyses

Cost-effectiveness plane

The bootstrapped ICER was assessed visually on the cost-effectiveness plane (Figure 12). Figure 12 shows that a large proportion of the bootstapped ICERs fall within the north-west quadrant meaning SLaMM+ intervention is more likely to generate less QALYs (be less effective) and incur more costs than SLaMM. Notably, some cost and QALY replicates fall within the other three quadrants, which demonstrates that there is uncertainty around the direction of the incremental costs and QALYs.

Reflection for item 14- What uncertainty analyses can be performed for trial-based economic evaluations?
Through this PhD, I became aware that the health economic literature recommends that uncertainty in the cost-effectiveness result is best represented through estimation approaches rather than hypothesis testing. Prior to commencing the PhD, I was familiar with representing uncertainty in my effects through 95% confidence intervals, nonetheless, I did not have experience of representing uncertainty in a ratio statistic such as the ICER. This required specialist knowledge and training in order to estimate uncertainty in the sampling distribution of the SLaMM+ and SLaMM participants and to present uncertainty through visual graphs such as the cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC). I provide a more detailed reflection on this item in the reflection box for item 14 within the Co-PARs study (Chapter 4).

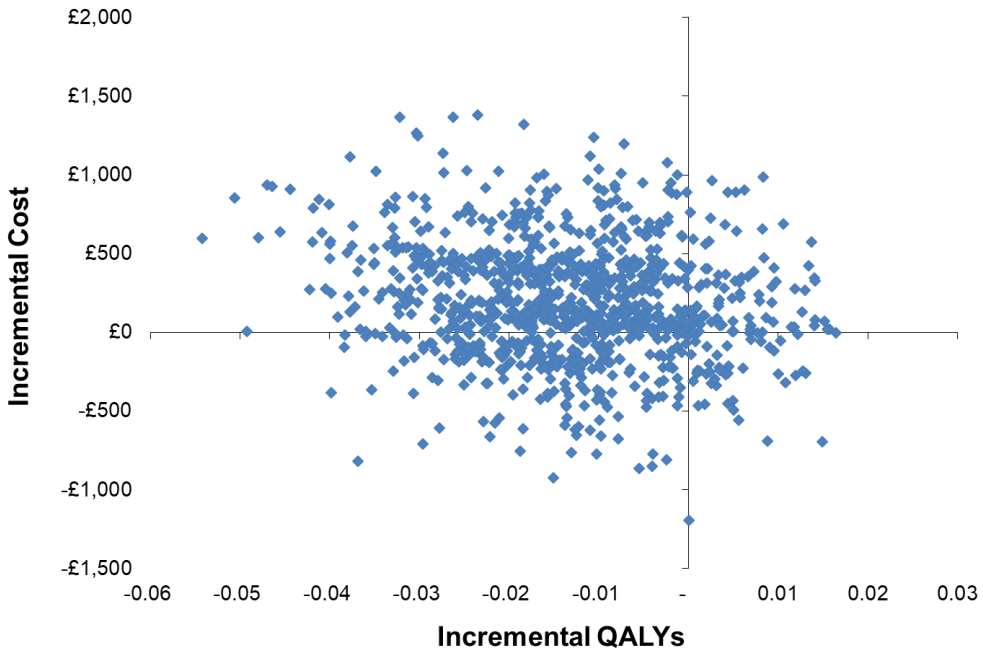


Figure 12. Cost-effectiveness plane for SLaMM+ vs SLaMM at 12 weeks

Cost-effectiveness acceptability curve

A cost-effectiveness acceptability curve (CEAC) was produced to help visualise uncertainty around the probability of being cost-effective. For example, Figures 13 show that the probability of SLaMM+ being cost-effective compared to SLaMM at a willingness-to-pay threshold varying from £0 to £50,000 per QALY. The curve shows that based on the participant-level data derived from this study and comparing with NICE's WTP threshold of £20,000-30,000 per QALY, there is only a 15% probability that SLaMM+ will be cost-effective compared to SLaMM.

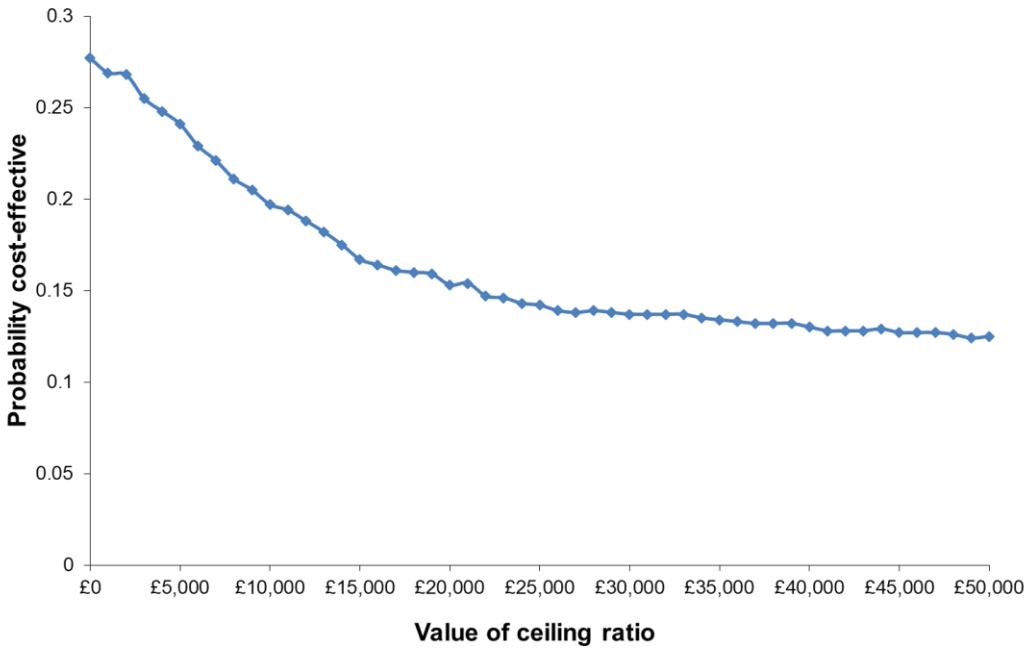


Figure 13. Cost-effectiveness acceptability curve showing the probability of short-term (at 12 weeks) cost-effectiveness for the SLaMM+ group vs SLaMM group at different willingness to pay per quality adjusted life year (QALY) thresholds

One-way scenario analysis

The aim of the one-way scenario analysis was to assess how sensitive the cost-effectiveness results were to the inclusion of the employer's intervention operating costs (employer's absenteeism costs were excluded due to this data being incomplete for 25% (n=10/40) of participants). This analysis may be of interest to public sector contact centres. The results show that the inclusion of employer's costs almost doubled the per person costs. As shown in Table 15, the results were the same as the base case results (differing by just £0.50), where under NICE's threshold, SLaMM+ was cost-ineffective compare to SLaMM.

Table 14. One-way scenario analysis for intervention costs including employer's costs

Variable	SLaMM+ per person	SLaMM per person
Base case: Public sector costs excluding employer's costs	£23.06	£8.83
Sensitivity analysis: Public sector costs including employer's costs	£40.51	£26.28
Base case ICER	£38,091.17 per QALY	
ICER adjusted for baseline HRQoL and healthcare utilisation	£38,091.67 per QALY	

5.4.2.3. Equity considerations

Equity subgroup analysis was not performed for the SLaMM+ group nor for the following subgroups for SLaMM: age, sex and deprivation as there would have been less than 10 observations per subgroup making it difficult to interpret these results (Table 16). As an alternative, equity was considered in the context of education level for the SLaMM group. As shown in Table 16, the analysis suggests that those educated to tertiary level are more likely to cost more but also experienced more QALYs. Under NICE's threshold of £20,000-30,000 per QALY, the ICER was not cost-effective and the incremental analyses showed the differences to be non-significant (Table 16).

Table 15. CUA for non-tertiary educated vs tertiary educated SLaMM participants

Analysis	Results
Incremental costs	Non-tertiary educated agents (n=11) cost £413.08 more per person, than tertiary educated agents (n=10)* (95% CIs: £-426.16 to £1,252.32; p-value=0.315)
Incremental QALYs	Non-tertiary educated agents (n=11) experienced 0.001 more QALYs per person, than tertiary educated agents (n=10)* (95% CIs: -0.010 to 0.012; p-value=0.860)
ICER	£413,080 per QALY

5.4.3. Cost consequence analysis

The CCA balance sheet (or impact inventory) is presented in Table 17 and provides a breakdown of the mean costs and consequences at 12 weeks (unless stated otherwise) from different perspectives: research institute, primary and secondary healthcare, employer, employee. Appendix D.4 provides a more detailed breakdown of the various costs and consequences. This balance sheet can be considered in conjunction with the primary outcome of the effectiveness evaluation, which has been reported elsewhere (Morris et al., in preparation).

Reflection on item 1.3- What components make up a well-defined study question? (Secondary Analysis)

The cost consequence analysis (CCA) lists the costs and effects relevant from the perspectives of the participants and the employer as the intervention takes place during employees' working hours. I felt that the presentation of the results in tabular form made it easier to discuss with the team and workplace how the two interventions differed in terms of delivery, set up, productivity and health costs and consequences. I found that discussing the cost categories with the most relevant stakeholders was easier and more appropriate than discussing a total cost which incorporated all costs (as shown in Table 14 of the CUA results).

Table 16. Cost consequence balance sheet

Costs			
Public sector perspective			
Variable	SLaMM+	SLaMM	Mean difference between groups (per participant)
Research Institute operating costs per person	£23.06	£8.83	+£14.23
Research Institute set up costs per person	£5.77	£3.36	+£2.41
Primary healthcare at 12-weeks per person*	£141.31	£50.76	+£90.56 (95% CIs:-£8.95 to £190.05; p-value=0.073)
Secondary healthcare at 12-weeks per person*	£355.08	£229.79	+£125.29 (95% CIs:-£576.19 to £846.757; p-value=0.727)
Private perspective			
Employer's operating costs per person**	£17.45	£17.45	£0
Employer's set up costs per person	£15.77	£9.26	+£6.51
Employer's absenteeism at Baseline (numerator/ denominator)	£24.74 (n=1/16)	£0 (n=14/14)	Insufficient data for comparative analysis

Employer's absenteeism at 12-weeks (numerator/ denominator)	£7.07 (n=1/16)	£14.81 (n=1/14)	Insufficient data for comparative analysis
Participant's private healthcare (numerator/ denominator)	£0 (n=19/19)	£0 (n=19/19)	£0
Participant's out of pocket (numerator/ denominator)	£0 (n=18/18)	£0 (n=19/19)	£0
Participant's loss in earnings at 12-weeks (uncertified sickness)*#	£92.52 (n=11/16)	£59.48 (n=10/14)	+£33.04 (95% CIs:-£21.17 to £87.24; p-value=0.222)*
	Consequences		
Public sector perspective			
Variable	SLaMM+	SLaMM	Mean difference between groups (per participant)
HRQoL at Baseline	0.780 (0.212)	0.884 (0.154)	-0.104 (95% CIs -0.222 to 0.136; p-value= 0.081)
HRQoL mean 12-week score*	0.882	0.830	
HRQoL mean change score*	0.047	-0.005	+0.052 (95% CIs:-0.041 to 0.146; p-value=0.266)*
Change in prescribed any medication	+20% (n=3/15)	+5.55% (n=1/18)	Insufficient data for comparative analysis
Change in moderate-to-strong painkillers	No change: 13.33% (n=2/15)	-5.56% (n=1/18)	Insufficient data for comparative analysis

Private perspective			
Variable	SLaMM+ % productivity loss	SLaMM % productivity loss	Mean difference between groups (per participant)
Employer's perspective: % productivity loss mean 12-week score*	19.48% (n=19/19)	18.63% (n=21/21)	+0.85% loss in productivity (95% CIs: -1.80% to 3.51%; p-value: 0.518)
Employee's perspective: Days of Uncertified sickness	1.18 (n=8/16)	0.64 (n=7/14)	+0.49 (95% CIs: -0.36 to 1.33; p-value: 0.249)

*adjusted for baseline value of that same variable; **assuming a 'steady state' by excluding set up costs;# managerial staff at the company reported that agency staff did not receive pay for uncertified sickness unless they claimed statutory sick pay.

5.4.3.1. Intervention costs

Table 17 presents the intervention operating and setting up costs the perspectives of the research institute and employer. From the research institute's perspective (a public sector agency) SLaMM+ costed £14.23 more per person than the SLaMM group. From the employer's perspective, there was no difference in operating costs between the groups. Mean set up costs were also greater in the SLaMM+ intervention compared to SLaMM. Appendix D.4 provides a more detailed breakdown of the intervention costs incurred by the research institute (payer) and the employer (setting). This breakdown makes it clear that after the height-adjustable desks, the second largest cost is staff time to attend three 30-minute education and training sessions.

Reflection on item 2 – What does a comprehensive description of the comparator groups look like?

It was planned that the trial protocol and CONSORT flow diagram would be consulted to describe the two comparator groups (SLaMM+ and SLaMM). From an economic perspective, the protocol and CONSORT flow diagram did not provide sufficient detail of all the resource types and quantities that are involved in the intervention. I felt that the microcosting exercise I conducted as part of item 6 in the framework generated a more comprehensive description of the intervention. Therefore, in future iterations of the framework I feel item 2 is not required as a separate item but can be incorporated into item 6.

Reflections on items 2 and 6.1– What does a comprehensive description of the comparator groups look like and how can intervention costs be measured?

I along with three Contact Centre staff provided data to populate the intervention microcosting database (Excel spreadsheets). The microcosting database comprised of data on the resource items and quantities required to set up and deliver the intervention. I was involved in the design, implementation and delivery of the SLaMM+ and SLaMM interventions and so I recorded this information on a weekly basis in an Excel spreadsheet. I also had informal conversations with the research team to verify that they agreed with the time costs I had allocated for certain intervention components. The aggregate data from the Excel spreadsheet are illustrated in Appendix D.2.

As discussed, I was directly involved in the delivery of the SLaMM+ and SLaMM interventions. This meant I was able to observe the complexity involved in implementing a new intervention into a workplace setting. One of the most important observations I made was that there were additional activities and time commitments being carried out by the staff at the workplace, namely the Resource Planner and two Centre Contacts, which I felt had not been quantified or captured yet in the intervention description sections of the protocol and CONSORT flow diagram. A further observation I made was that the Centre Contract felt it was necessary to have two Centre Contracts rather than one in order to support the implementation and delivery of the interventions. I believe my Masters in Public Health training informed my decision to capture these modifications and understand the complexity involved in implementing an intervention. During my Masters I learnt about a key challenge faced by the health community called 'the implementation gap'. This is the idea that a large proportion of findings from high-quality research studies fail to be implemented into the 'real world' due to the research studies not capturing the complexity involved in implementing evidence into practice (Haines et al., 2004).

In order to capture the additional activities and time commitments of the two Centre Contracts and Resource Planner, I requested that these three staff complete a weekly electronic log over the 12 week intervention period. In my initial framework, I had not recommended this approach as I thought it would suffice to capture these additional costs retrospectively through telephone interventions. However, I was aware that the organisation we were working with has a high turnover of staff and so I felt it was best to collect this data prospectively. I therefore did some additional literature searching to see if I could find a template to use for the staff to log their activities and time commitment. I came across a public health study which had used weekly diaries with intervention staff for their microcosting exercise (Charles et al., 2013). I used the structure of these diaries to inform my own electronic logs. Nevertheless, completeness of the electronic logs was poor for the two Centre Contacts. I tried several strategies to support the two Centre Contracts to complete their logs including sending email and telephone reminders at the

start and end of each week for three consecutive weeks. Nonetheless, the Centre Contracts reported not having the time to do this on a weekly basis. The Centre Contracts agreed to estimate the time they spent implementing the intervention retrospectively through a one-hour face to face informal meeting with me at the 12-week follow up time point. As the two centre contacts preferred to meet face-to-face to discuss intervention resource use with me, this made me aware that the approach used to collect data for the microcosting exercise requires some degree of flexibility. By contrast, the Resource Planner reported no issues with completing electronic weekly logs.

In the initial framework, I designed and presented a non-study specific interview schedule and microcosting tool which could be used to capture intervention resource quantities. Through my observations during the trial, I found my interview schedule and microcosting tool was too simple and would not capture all the complex resource use involved in the delivery of a complex workplace trial (SLaMM). I therefore added content to the weekly log and informal meeting schedule (see Appendix D.2). This content was informed by a tool from the public health and behavioural science literature, the TIDieR framework (Hoffmann et al., 2014).

Reflections on item 8.1- How can costs be valued? (Intervention costs)

The reflections I describe in the reflections box for item 8.1 in Chapter 4 for the Co-PARs trial also apply here to the SLaMM trial. For example, costing the intervention involved numerous calculations and the assumptions. These calculations were recorded in an Excel spreadsheet (see Appendix D.2). A budget breakdown was not available for the intervention costs since the funding application for the SLaMM trial related to the funding of two PhD students' and one MPhil student's research programmes rather than an independently funded trial. I used national published costs for the salaries of the staff from the company receiving the SLaMM trial. Nevertheless, calculating the unit costs was a time-consuming task as it involved considering overhead costs, as well as the intervention-specific capital costs. For example, in order to assign a unit cost to the height-adjustable desk required for the SLaMM+ intervention I needed to make several adjustments to the price reported in the expense claim forms. This was because I had read in the health economic literature that assigning a unit cost to capital equipment, which would not have been purchased on the same scale without the intervention, requires consideration of the "life" of the equipment (Drummond et al., 2015a). I followed the guidance by Drummond et al. (2015) and assumed that the life of the desk would be 5 years. From here, I calculated how much the height-adjustable desk cost for the 12-week time horizon of the trial. For other capital, such as the rooms used for the education and training sessions, I assumed the opportunity cost was small and so did not assign a unit cost to these resource items. I felt it was appropriate to assume the opportunity cost was small since the rooms and IT systems had pre-existed at the company prior to the SLaMM trial being introduced. The costing procedures described here, illustrate the complexity involved in costing and how I found it difficult to recommend a standardised approach that would not be influenced by the study's context.

5.4.3.2. Healthcare utilisation

At 12 weeks the SLaMM+ group had incurred greater primary and secondary care costs compared to the SLaMM group (Table 17). Nonetheless, as the 95% confidence intervals suggest, there is a lot of uncertainty in these results and the mean difference was not statistically significant (Table 17). Appendix D.4 provides a breakdown of the most commonly used and most costly primary and secondary care activities. Out of all healthcare activity, the most commonly used service was the GP, followed by the Practice Nurse and Counsellor. Although, secondary care activity had the greatest unit costs, at both time points secondary care was only utilised by a minority of participants.

Reflections on item 8.2- How can costs be valued? (Health and social care costs)
Assigning unit costs to the healthcare activity reported in the questionnaires involved several calculations and assumptions. These were recorded in an Excel spreadsheet for transparency (see Appendix D.3). The reflections reported in the reflection box for item 8.2 in Chapter 4 (the Co-PARs study) apply to the SLaMM trial, as the trials were conducted concurrently.

5.4.3.3. Prescribed medication

In both groups, medication use did not decrease over the 12-week period. The type of medications taken varied widely. Across both groups, moderate-to-strong painkillers were the most commonly prescribed medication (Table 17). A more detailed breakdown of the most common medications is provided in Appendix D.4. Importantly, 17.5% (n=7/40) participants did not report their medication use.

5.4.3.4. Employer's productivity loss

Employer's costs was only included for participants who were off sick but the sickness was certified (approved by a medical professional). For certified sickness, it was assumed Statutory Sick Pay (SSP) would be paid to the agent by the employer. In both comparator groups, just one participant was driving the certified sickness estimates at 12 weeks. Nonetheless, it should be noted that this data was only available for 75% (n=30/40) of participants. A minority of participants took uncertified sickness for four or more days in row (Appendix D.4).

Reflections on item 8.4- How can costs be valued? (Productivity costs)
As uncertified sickness was costed from the participant's perspective (see reflection boxes for items 4 and 8.3), only certified sickness was costed from the employer's perspective. A human capital approach was used which may have overestimated these costs (Hanly, 2012). Nevertheless, just one participant per comparator group at each time point reported a certified sickness day, making the likely impact on the results small. I had seen that an alternative approach to calculating absenteeism costs called the friction approach, was reported in the literature. However due to the short follow up period of this study, I felt it was more appropriate to carry out a human capital approach. In terms of assigning a unit cost to presenteeism, I could not find any practical guidance in the literature on how this could be done. I sought advice about this during my study placement at Deakin University. There was consensus between the health economics group that due to the subjective nature of presenteeism, it was difficult to assign a monetary value to presenteeism. As an alternative, I believed it was appropriate to report this data in its natural units.

5.4.3.5. Participant costs including loss in earnings

No participants reported incurring any out of pocket expenses as a result of participating in the intervention. Mean loss of earnings, based on the amount of time taken as uncertified sick leave, was a relevant cost that nearly all participants incurred during the 12 weeks (Table 17). The

difference in mean loss in earnings between the groups was uncertain based on the 95% CIs and p-value.

5.4.3.6. Health-related quality of life (HRQoL)

Health-related quality of life (HRQoL) as measured with the EQ-5D-5L was greater in the SLaMM group at baseline, but not significantly greater (Table 17). At 12 weeks the SLaMM+ group had a greater HRQoL score (+0.052), but the 95% CIs demonstrated this difference was not statistically significant.

Reflections on item 15- How can the results be interpreted?

I reflected on my analyst-based decisions in relation to the design of the SLaMM trial in order to consider how these design decisions relate to the applicability of my results to similar workplaces who are interested in implementing a SB intervention. A tool for assessing applicability of trial-based results is the PRECIS-2 tool. I discuss the PRECIS-2 tool in greater depth in reflection box for item 3 for the Co-PARs trial (Chapter 4). In brief, the tool lists nine trial design features which can be reflected on in order to consider whether the trial is more pragmatic or explanatory orientated. In the initial framework in Chapter 3, I recommended reflecting on the setting in which the trial is conducted but I had not previously considered other design features which can improve the applicability of trial results to the 'real world'. These design features from the PRECIS-2 tool include: delivery of the intervention in a way that could be done if rolled out across multiple workplaces; encouraging participants to adhere to the intervention in a similar way that would be possible in the 'real world'; following up patients to assess the benefits of the intervention to the similar intensity that would be done if the intervention was rolled out in practice; using an outcome measure that is important to the participants; enrolling participants onto the intervention in a similar manner that would be done if implemented without the research team as well as ensuring these participants are similar to those who would be eligible if rolled out; and conducting an intention-to-treat analysis. I believe my economic perspective helped me recognise that if an intervention is not delivered in the same way as it would be in the real world, then this means the resource use data I was presenting would not be useful for 'real world' decisions.

Prior to conducting the SLaMM trial I had limited experience of designing a RCT in order to support decision-making. In part, my background in anthropology has led me to believe the PRECIS-2 tool is important. This is because anthropology is about describing complexity, which I feel the PRECIS-2 tool also aims to capture through the nine domains of trial applicability. In addition, my key hands-on role in the setting up and delivery of the SLaMM+ and SLaMM interventions also made me aware of the amount of resource (namely organisation time and travel time) that is involved in the implementation of interventions. I feel these time costs represent hidden costs which are not usually documented in published effectiveness evaluations but may contribute to the 'implementation gap' phenomenon (see the reflection box for item 3 in Chapter 4 for further discussion about the 'implementation gap').

Another key aspect that attracted me to the PRECIS-2 tool is its focus on making it explicit what the intended purpose of the trial. A key feature of a good quality economic evaluation is about being explicit about what the intended purpose of the trial by reporting the study perspective (perspective is more commonly referred to as the 'stakeholders' in the public health literature). Clinical effectiveness trials do not typically report who their study is aimed at but the inclusion of this factor in the PRECIS-2 tool suggests that there might be a move towards requesting trialists to be more explicit about who their results intend to inform.

Reflections on item 16- how can trial-based economic evaluations be reported?
I found the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist useful for writing up the methods and results for this study. Nonetheless, during the first six months of this PhD project (late 2016-early 2017) when I was still relatively new to the discipline of health economics, I found the terminology in the CHEERS checklist challenging to understand as there are many concepts which are not used in the effectiveness evaluation literature. I feel therefore feel that the terminology used in CHEERS may therefore act as a barrier to researchers or decision-makers without health economic training. This is an issue if it prevents: (1) researchers from adequately reporting their methods; and (2) decision-makers from using the results from economic evaluations.

5.5. Discussion

The overarching aim of this study was to apply the initial framework from Chapter 3 to the SLaMM trial. More specifically, the study had two concurrent aims: (1) to present an economic evaluations of the SLaMM trial; (2) to reflect on the applicability of the framework to a real world SB trial. The discussion therefore starts by an interpretation of the results of the economic evaluation, before going on to discuss the reflections from the initial version of the framework and consider how the framework could be improved for future application.

5.5.1. Principle findings: CUA

The primary economic analysis of this study was to evaluate the cost-effectiveness of the SLaMM+ intervention using CUA. The comparative analysis of the SLaMM+ intervention to the SLaMM intervention generated a cost-effectiveness ratio of £38,091 per QALY. This ICER is higher than NICE's maximum willingness to pay threshold of £20,000-£30,000 per QALY. This suggests that SLaMM+ is not cost-effective compared to SLaMM. That said, the results of this analysis should be treated with caution. The standard deviations, 95% confidence intervals, cost-effectiveness plane and cost-effectiveness acceptability curve indicate that there is a lot of uncertainty in these results. In particular, uncertainty can be seen in the shape of the CEAC, which shows that as a decision-maker's willingness to pay threshold increases, the probability of SLaMM+ being cost-effectiveness compared to SLaMM decreases. This suggests that the cost-effectiveness results may be driven by the between group difference in QALYs (EQ-5D score). Overall, the summary ICER result is not supported by the uncertainty results. In part, the difference may reflect the type of analysis techniques selected. For instance, the nonparametric bootstrapping simulation draws on original cost and QALY data. By contrast, the ICER statistic has been adjusted for imbalances in costs and QALYs using multiple regression, as recommended in the literature (Franklin et al., 2019, Manca et al., 2005).

5.5.2. Comparison with other studies

A similar trial-based economic evaluation from Australia which assessed a SB workplace intervention called Stand Up Victoria (Gao et al., 2018) was identified in Chapter 2's systematic review. It is challenging to compare the results of the present study with the study by Gao et al. (2018) since the authors do not report their results as 'ICER per quality of life year'. Instead, the analysts present 'ICER per reduction in sitting time'. In the UK, the ICER per QALY is the preferred way for summarising economic evaluations (NICE, 2014a). In terms of presenteeism, the present study found no significant difference in presenteeism at 12 weeks. A similar UK-based RCT called the Stand More at Work (SMArT) trial, also observed no differences in presenteeism at 3 months (Edwardson et al., 2018). On the contrary, presenteeism was not included nor commented on the Stand Up Victoria study (Gao et al., 2018). Instead, the authors restricted their analysis to absenteeism at 12 months, for which they identified a 2-day difference in sick days between the comparator groups. Alike the difference in uncertified sickness in this study, the results were not statistically significant. No significant difference in HRQoL was observed between the SLaMM+ and SLaMM agents at 12 weeks. Similarly, the trial by Gao et al. (2018) found no significant difference in quality of life in their trial-based analysis. These findings contradict the results of the SMArT trial by Edwardson et al. (2018) which did observe a significant increase in quality of life in the intervention group compared to the control group (usual practice) at both 6 and 12 months. A recent high quality meta-analysis of 18 original studies found lower levels of SB are associated with higher quality of life indicating this is an important outcome measure to assess (Boberska et al., 2018).

There are key methodological differences between the SLaMM, SMArT and Stand Up Victoria analyses, which may explain some of the inconsistencies in results. For example, all three trials used different quality of life measures as well as different intervention components. Furthermore, the control groups for the SMArT and Stand Up Victoria trials received no intervention components (only results of their health measures in the SMArT trial's control participants) while the SLaMM comparator group received multiple intervention components except for a height-adjustable desk. Secondly, the SMArT and Stand Up Victoria were larger trials with longer time horizons, as 107 and 167 participants provided 12-month follow up data respectively. Nevertheless, a key strength of the SLaMM trial is that out of the three, it was the only study which presented a comparative analysis of healthcare utilisation. This is surprising since the systematic review in Chapter 2 found that comparative analyses for primary and secondary healthcare use was performed in the majority of PA studies. Despite this study having low power, the evidence in this trial contributes to the literature since there is lack of evidence on the cost-effectiveness of individual-level SB interventions (Chapter 2).

5.5.3. Principle findings: Application of the framework recommendations

Table 18 shows that the majority (n=14/ 16) of the items in the initial version of the framework require some form of modification (structural or additional guidance). More specifically, 11 items were not easy to implement and required analyst-based decisions to be made. For these items it was felt additional guidance will be needed to ensure the framework facilitates a standardise approach. Recommendations on how the framework could be improved future application are

outlined below. These recommendations have been based on the reflections in Table 18. Framework items in which the recommendations relate to are signposted in parentheses.

5.5.3.1. Primary economic analysis

Similar to findings in Chapter 4, additional guidance is required to support analysts conducting economic evaluations alongside pilot trials. The refined framework should recommended that the provision of a breakdown of costs and consequences the primary analysis (see reflections in item 1.2 and 1.3). Summary of the costs and consequences should be presented by sector in tabular form, as seen in Table 17 in this study. This disaggregated format has been used in existing high quality studies (Jacklin, 2003) and is recommended as the results are easier to understand. The research should also point out to the reader that this disaggregated cost and consequence data can be used to inform the design of future larger trials and interventions (see reflections in item 1.3). In summary, CUA as well as decision-analytical modelling is not recommended as the primary analysis for the framework (see reflections in 1.4 and 7.2).

5.5.3.2. Trial design

Similar to findings in Chapter 4, the refined framework should recommend that the economic researcher is involved in the trial design from the outset to ensure the trial is fit for purpose (see reflections in item 3 and 15). External validity is important for economic evaluations as well as behaviour change interventions. Historically, within the general literature for health research there has been a lack of focus on the external validity of trials (Rothwell, 2005). The nine domains of a pragmatic trial described in the PRECIS-2 tool can be drawn upon (Loudon et al., 2015) as demonstrated in section 5.4.3.2.

5.5.3.3. Incorporating productivity

All 30 agents with complete absenteeism data in this study were agency workers. Reportedly, this meant they were not eligible for statutory/ company sick pay unless they were absent for 4 or more days in a row. As most agents in this study took less than 4 days sick days in a row, uncertified absenteeism data was costed from the participant's perspective (see reflections in 6.4). The framework should therefore recommended that contextual factors related to the employees' contract is captured. Importantly, in the UK, if the sick employee is sick for less than 4 days and is contracted via an agency, then they may be the one who experiences a loss in earnings, as opposed to the employer (HM Treasury, 2019). For this scenario, the framework should recommend using the human capital costing approach, which takes the perspective of the sick employee, when estimating productivity losses.

5.5.3.4. Cost drivers for future data collection

In the pilot study the economic data collected can be simplified to the focus on collecting data for the key cost drivers. In terms of intervention costs, this study found that staff time was a cost driver for the intervention set up and operating costs from the perspectives of the research institute and employer. For example, the Resource Planner's time to schedule the Education and Training sessions, had not previous been reported as a key activity in the trial protocol and CONSORT diagram (see reflections in items 4.1 and 6.1). Similar to the recommendations in Chapter 4, the framework can recommend that the TIDieR checklist (Hoffmann et al., 2014) and mircocosting tool

by Charles et al. (Charles et al., 2013) can be used to structure data collection methods such as consultations and weekly electronic logs, respectively. Key cost drivers from the healthcare perspective that the framework can recommend as priority items for data collection include: all secondary care activities and primary care consultations with the GP, Practice Nurse and Counsellor (see reflections in item 4.2). No participants reported incurring any out of pocket costs due to participating in the intervention. This is likely to be due to the aim of the intervention only requiring agents to predominantly stand and walk more during work hours. The framework can recommend that this type of cost does not need to be prioritised for data collection in similar workplace studies, if research resources are restricted (see reflections in item 6.3).

5.5.3.5. Assigning unit costs

Similar to findings in Chapter 4, the framework should recommend that all capital equipment needs to be quantified and reported alongside the other intervention resources. However, it can be suggested that not all items may need to be assigned a unit cost. Overhead and capital equipment costs can be excluded, if the resource (e.g. private room/ lighting) would not have been used for an alternative purpose (opportunity cost) and the resource use is minimal. This approach has been applied within the public health economics literature (Edwards and McIntosh, 2019). Capital equipment required only for the intervention (e.g. the height-adjustable desk) should be assigned a unit cost and included in the total intervention cost calculation. If the life of the capital equipment being costed is unknown, 5 years can be used, as this is seen as a conservative estimate (Drummond et al., 2015a). Furthermore, the framework can advise that the unit cost estimate for key capital equipment is adjusted according to the number of weeks the equipment is required for. This approach was used by Anokye et al. (2018) in a recent PA trial (see reflection in 8.1). In relation to unit costs, the framework can recommend that until more standardised methods are developed for measuring and valuing medication use (Thorn et al., 2018), this data does not need to be prioritised for data collection in SB trials, or if deemed important for a specific study, it can be reported descriptively (see reflections in 8.2). Lastly, the framework can provide additional guidance on costing productivity. For example, it can be advised that if productivity is interpreted as loss of earnings from the employee's perspective, the human capital approach is appropriate. Similarly, this approach is also deemed acceptable for small scale studies, since it is easier to apply and commonly used (Edwards and McIntosh, 2019). On the contrary, larger studies exploring productivity from the employer's perspective should aim to use a friction approach since the human capital approach is expected to overestimate productivity loss in this context (van den Hout, 2010) (see reflections in 8.3 and 8.4).

5.5.4. Strengths and limitations

Strengths and limitations of the trial design have been reflected on firstly by considering the internal validity of the results drawing on concepts from the CONSORT framework for social and psychological interventions (Grant et al., 2018). Secondly, the PRECIS-2 tool (Loudon et al., 2015) has been used to assess the external validity (generalisability) of the study. PRECIS-2 is a well-known tool developed by 25 international trialists and methodologists, with the aim of providing a tool which support researchers to reflect on the design of their trials in a systematic, transparent and comprehensive manner. The tool aims to prevent waste in health research.

5.5.4.1. Internal validity

A strength of this trial has been the random allocation method used to assign participants to the two comparator groups. For example, the demographic data indicates that randomisation controlled for a number of demographic variables. Nevertheless, even when randomisation is used, the health economic literature recommends that baseline costs and HRQoL are controlled for when estimating total costs and QALYs, as was done in this study (Manca et al., 2005, Franklin et al., 2019). The multiple regression method used in this study has therefore enabled more precise estimates to be generated. That said, even though baseline imbalances were adjusted for, no statistically significant differences were observed between the groups for any of the costs or consequences measured. This may be due to there being insufficient power to detect a difference and/ or the time horizon of the study being too short. One way to conduct addressing the missing data could have been increased, would have been by performing multiple imputation methods rather than a complete case analysis. There is a lack of guidance on how to perform multiple imputation in trial-based economic evaluations which has been highlighted in the literature (Leurent et al., 2018) meaning complete case is the most common approach used for handling missing data in economic evaluations (Faria et al. 2014). Similarly, low power has been outlined as being a common issue within this field of research. A review from 2016 found that most effectiveness evaluations which assessed workplace height-adjustable desk interventions had low power and high risk of bias (Shrestha et al., 2016). Due to the small sample size, the pre-specified equity impact analyses for age, sex and postcode were not performed. An additional limitation related to the data available, was that there was incomplete absenteeism and medication data, 25% (n=10/40) and 17.5% (n=7/40) of data were missing respectively.

5.5.4.2. External validity

The generalisability of the SLaMM and SLaMM+ interventions to other workplaces is unknown since the study was conducted in just one workplace. That said, the agents who took part in the trial are likely to be representative of the contact centre since the eligibility items was kept intentionally broad. A limitation of this study is that the per-protocol principle was whereby participants were excluded from the study if they did not adhere to the intervention they were assigned to. Per-protocol principle restricts the readers understanding on how resources are actually being used in practice (Ramsey et al., 2015). Some aspects of the interventions delivery are unlikely to reflect what would happen if the intervention was scaled up, this includes the education and training sessions being delivered by the research team. Measuring the healthcare utilisation, presenteeism and HRQoL variables through self-report questionnaires was cheap and easy. In this trial, these measures were collected face-to-face by the research team, however future studies could collect this data in practice with little resource impact from the employer's perspective. This may be especially true if the questionnaire is disseminated and completed via a secure confidential link online. For example, the Centre Contact could disseminate the link via email, similar to how the weekly emails were disseminated, which reportedly took no more than 5 minutes to do. A strength of using this method would be that the participants who had left the company and thus did not complete the 12-week follow up could be also sent this link if they consented to this approach at the start of the trial. Inevitably, self-report measures are subject to

bias, however subjective measures are the preferred approach for presenteeism and healthcare utilisation due to objective data being incomplete, difficult to obtain and having poor external validity (Franklin and Thorn, 2019, Kigozi et al., 2017).

5.5.4.3. Economic studies alongside pilot RCTs

One of the main limitations of this study was that the economic evaluation was carried out alongside a pilot RCT. This meant the sample size was small and the study did not need statistical power for the clinical outcome. Furthermore, the sample size reduced further due to the contact centre company experiencing a high-turnover in staff. As this study was a pilot RCT it was the first time the trial team had worked with this type of contact centre company. The team had therefore not anticipated that a fifth (20%, n=12/60) of participants would leave the company during the 12-week trial period and therefore discontinue with the study. No data was collected for these participants at the 12 week period and so we do not know for certain whether they experienced different outcomes to those who remained in the study. Another consideration with a pilot RCT and small sample size, is that there is a greater risk of outliers having a dramatic impact on the cost data (Simpson et al., 2014). Nevertheless, some argue that cost data is typically skewed anyway and has typically has a variance than clinical outcomes (Briggs, 2000).

The advantage of conducting an economic evaluation alongside a pilot RCT is that it provides the opportunity to collect some evidence on the costs and implications of the intervention which may be important if there is not the time or funding to conduct a full trial (Glick et al., 2014). Collecting data alongside small scale studies may be particularly relevant to public health decision-makers in the UK who operate on short-term annual budgets and therefore require evidence to be generated rapidly (Willmott et al., 2015). Another advantage of incorporating economic procedures in a pilot trial is that it allows the trial team to assess the feasibility and acceptability of the data collection methods (Blatch-Jones et al., 2018) and refine their methods for future trials, as is the case with my framework. Moreover, an additional argument for collecting economic evidence alongside a pilot RCT is that even within full RCTs the statistical power for the economic end point is rarely calculated. In part this is because of the complex nature of the outcome of interest in economic evaluations (cost-effectiveness result) which is the joint distribution of differences in costs and effects (Petrou and Gray, 2011). Furthermore, due to the large variability in cost data, it is expected that a very large sample size would be required which may be unethical and costly. For these reasons, some argue that the focus of economic evaluations should be on reporting the cost-effectiveness result because the confidence intervals will reveal if a study has low power (Briggs, 2000). The wide confidence intervals for the cost data presented in this present study (SLaMM trial) indicate this study was underpowered.

5.6. Conclusion

In order to fulfil the first aim of providing a multidisciplinary framework that ensures a standardised approach to conducting economic evaluations, there is a need to improve costing methodology in particular, the methods used for capturing productivity and assigning unit costs. Importantly, the key reflections based on piloting of the initial version of the framework to this empirical trial have highlighted how the framework can be improved in order to emphasise the importance of

disaggregating costs and consequences and by collecting contextual data in order to consider who the results may be relevant to. The second aim of the study was to present the results of the CUA. The ICER statistic indicated that at 12 weeks when SLaMM+ was compared to SLaMM, the SLaMM+ group had an ICER that was just over NICE's willingness to pay per QALY threshold. The uncertainty analyses indicated that there is a low probability the SLaMM+ will be cost-effective compared to SLaMM. Though, this study was a pilot RCT and thus was underpowered to make more robust conclusions. The second aim of the study was to reflect on the application of the initial version of the framework to the SLaMM trial. The framework was not straightforward implement as around two thirds (n=11/16) of the items needed additional guidance.

Chapter 6: Synthesis of findings & presentation of a standard operating procedure

6.1. Introduction

6.1.1. Revisiting the aim & objectives

The overarching aim of the PhD was to develop a multidisciplinary and pragmatic framework to guide researchers in conducting economic evaluations of individual-level PA and SB interventions. In order to address this aim, the following three objectives were undertaken:

1. To conduct a systematic review of the empirical literature in order to identify and critically appraise the methods applied in existing economic evaluations of individual-level PA and SB interventions (Chapter 2, Study 1).
2. The development of an initial multidisciplinary and pragmatic framework, which draw on good quality methods identified from Study 1 (Chapter 3).
3. To apply the initial version of the framework to one PA trial (Chapter 4, Study 2) and one SB trial (Chapter 5, Study 3) in order to report and reflect on the applicability of the framework as well as provide evidence on the cost-effectiveness of the interventions under study.

Section 6.2 provides an overview of each chapter's methods and contributions. In section 6.3 I discuss the narrative synthesis methods I used to merge the reflective content from the thesis. Section 6.4 presents an overview of key findings from the narrative synthesis which relate to: (1) the challenges I came across when implementing the framework; and (2) the actions I took to facilitate implementation of the framework. In sections 6.5 I provide a critical discussion of these challenges and facilitators by interpreting them in the context of the wider literature. In section 6.6 I outline the implications for research practice. Lastly, in section 6.7 I present a revised version of the framework in the form of a standard operating procedure (SOP). In research, a SOP is guidance document which outlines a clinical trial unit's preferred methods for carrying out a specific procedure. In brief, SOPs play an important role in quality assurance, transparency and consistency (Dritsaki et al., 2018), similar to the role I believe my framework can play. In section 6.6.4, I explain further why I have chosen to disseminate my framework in the form of a SOP.

6.2. Overview of thesis chapters

6.2.1. Chapter 2: Systematic review

The aim of the systematic review was to understand how existing economic evaluations of individual-level PA and SB have addressed four key methodological challenges. Just 15 studies were identified from four countries. Furthermore, no interventions being evaluated in these economic evaluations had assessed SB as an independent risk from PA, before 2018. Just one study made assumptions around the long-term impact of the PA intervention which was supported by evidence. This demonstrated that progress is being made to address the challenge of linking up short-term and long-term evidence. Though many country's provide guidelines on their preferred techniques for measuring and valuing outcomes in economic evaluations, there was inconsistencies within studies from the same country on how resources were measured and valued. Comparability of studies was therefore challenging. Furthermore, the literature had

suggested that outcomes may be valued using CBA technique, but the review identified no examples of this. The most frequent costs and consequences incorporated into the economic evaluations reviewed were intervention operating costs and healthcare utilisation costs. A minority of studies included a slightly broader range of costs including: participant's out-of-pocket costs; participant's productivity loss (time off work to participate in PA intervention); provider's productivity costs (loss in revenue); and provider's intervention set up costs. The systematic review found that all but one study incorporated equity into their study, demonstrating that this is common practice. That said, only one study looked at more than one equity subgroup, which suggested that equity was not being considered in an explicit nor systematic way.

6.2.2. Chapter 3: Development of an initial framework

Four key steps were taken to develop the initial version of the standardised framework. The first step involved specifying the structure for the framework. The structure included 16 items which were deemed important and were predominately informed by the 10 items in the methodological quality assessment checklist by Drummond et al. (2015). In addition, the structure aimed to be incorporated the learnings from my systematic review (Chapter 2) such as a greater emphasis on incorporating equity and outlining the identification, measurement and valuation procedures separately for costs and effects. The second step involved drawing on the findings from my systematic review in terms of what costs, outcomes and equity subgroups are appropriate to recommend in the framework and in order to address the methodological challenges associated with public health economics. The third step entailed identifying data collection tools to measure the costs and outcomes deemed relevant. This step was challenging, since by large the studies from my systematic review did not report or provide examples of the measurement tools they had used to collect healthcare utilisation, intervention resource use and productivity data. Consequently, the DIRUM database was used to identify tools from other disease areas, which could be modified. Lastly, the fourth step involved arranging regular informal meetings with my supervisory team, trial team and other experts in order to discuss and reflect on the relevance, importance and practicality of the approach I was specifying in the framework and to ensure the approach was multidisciplinary and pragmatic. A key learning in the development of the initial framework was that standardisation of some methodological steps was constricted due to the need to recommend that the reader consults national guidelines, costing databases and/or position statements for the country in which they are conducting the analysis from.

6.2.3. Chapter 4: Application of the framework to the Co-PARS trial

The overarching aim of this study was to pilot and test the initial framework in a 'real world' PA trial. The study had two aims which were to reflect on the application of the initial version of the framework, and to assess the relative cost-effectiveness of the Co-PARS intervention compared to usual care and the control group. An important finding from the reflections on applying the framework was that it was challenging to implement many items of the framework without practical and formal training in economic evaluation. Furthermore, my reflections on the ICER result, led me to believe that the ICER statistic generated from a CUA is too simple and hides the complexity and wider benefits and costs of delivering a PA intervention such as Co-PARs. By contrast, I found that my secondary analysis, the cost consequence balance sheet where costs were disaggregated

(Appendix C.4), provided a whole range of information relevant to a number of different stakeholders (perspectives) including the participants. It was felt that this data was more helpful than the ICER for a number of reasons: the disaggregated results helped in me understand what the key cost drivers are (e.g. secondary healthcare costs and GP visits). In addition, my reflections illuminated the practical challenges in costing methodology in terms of: (1) how to assign unit costs in a standardised way without the need to make additional analyst-based assumptions and calculations; and (2) how to simplify the measurement of participant's out-of-pocket costs and prescribed medication costs. More specifically, the unit cost calculations presented in Appendix C.3 demonstrate the various levels of precision, adjustments and secondary sources required for costing. Finally, it was not possible to produce a decision-analytic model due to there being insufficient time, expertise and robust data from the trial. In terms of cost-effectiveness, the ICER indicated that from NICE's perspective, the Co-PARS intervention was cost-effective compared to usual care but not compared to the control group. Though these results should be treated with caution due to the small study size. A future study with larger power is recommended to better understand the probability of the Co-PARs intervention being cost-effective compared to usual care. The stochastic uncertainty analyses indicated there was a lot of uncertainty in the original Co-PARs dataset.

6.2.4. Chapter 5: Application of the framework to the SLaMM trial

The overarching aim of this study was to pilot and test the initial framework in a 'real world' SB trial. The study had two aims, which were to reflect on the application of the initial version of the framework, and to assess the relative cost-effectiveness of the SLaMM+ (height-adjustable desk group) intervention compared to the same intervention without a height-adjustable desk. My reflections on applying the initial framework to the SLaMM trial build on the findings observed in the Co-PARS study, specifically in relation to the usefulness of the ICER summary statistic in the SLaMM trial. Nonetheless, there were some differences that predominately related to the setting in which the framework was implemented. The SLaMM trial was delivered in a workplace which provided an opportunity for me to explore the employer's perspective and access objective absenteeism data. That said, the absenteeism data precluded a comparative analysis, since a large proportion of this data was incomplete. I believe, this highlights the need to understand the quality of the existing datasets that may be available, before the trial commences, so as solutions can be sought to improve the quality of the collection recorded prospectively. The SLaMM trial helped me realise the importance of checking my assumptions around how the company hosting the intervention is organised, in order to avoid assigning resource use and savings to a narrow or incorrect perspective. For instance, prior to informally discussing how the company was organised with the senior managers, I had assumed that all paid work productivity loss would be incurred by the employer. However, I discovered it was actually the employees who were also likely to have a loss in earnings if they were off sick due to the nature of their job contracts. An important finding from the SLaMM trial, which differed to the design of the Co-PARs trial was that the researchers (including myself) delivered a large proportion of the intervention. By participating in the intervention delivery, I was able to provide prospective data on the type and amount of resources used for the intervention set up and operating costs. Nevertheless, on some weeks, some

company staff reported inputting more time than expected into supporting the implementation of the intervention. I asked these staff to document their time by completing weekly logs. For two of the three company staff, they reported not having the time to complete the logs on a regular basis. This showed that flexibility is needed for some key informants, for instance, this data may need to be collected retrospectively if weekly logs or diaries are deemed to be too onerous. In terms of cost-effectiveness, the ICER indicated that from NICE's perspective, the SLaMM+ intervention was just over the cost-effectiveness threshold. The small sample size as well as the results of the uncertainty analysis make it challenging to make conclusions about the cost-effectiveness results.

6.3. Methods of narrative synthesis

The reflective content interwoven throughout this thesis was based on a variety of ethnographic data sources, primarily these included the observations, notes and informal conversations which surfaced during the development and application of my framework. As a result, the reflections provide valuable insight into of the actions I took in order to develop and apply the framework in practice. Importantly, the reflections also provide explanations for why I believed these actions were appropriate. In order to revise and improve the initial framework from chapter 3, I have used methods of narrative synthesis to merge and organise the reflective content from chapters 4 and 5. The narrative synthesis involved organising my reflections into the two broad themes which relate to the aim of my framework, multidisciplinary and pragmatism. Interpreting the multidisciplinary and pragmatic content of my framework helped me discover that my reflections relate to two more specific themes: (1) the practical challenges a researcher may face when conducting an economic evaluation of an individual-level PA and SB intervention; and (2) the actions a researcher could take in order to support the implementation of an economic evaluation of individual-level PA and SB interventions. I provide a critical discussion of the factors which relate to these themes by interpreting them in the context of literature from health economics, public health, behavioural science and trial methodology.

6.4. Narrative synthesis of key themes

6.4.1. Overview of key challenges to framework implementation

6.4.1.1. Complexity

The reflective content in this thesis highlighted the complexity in developing and applying a standardised and principle-based framework for the economic evaluation of individual-level PA and SB interventions. One reason for this complexity relates to my observation that economic evaluations of individual-level PA and SB interventions requires skills and knowledge that go beyond the traditional disciplines of evidence-based medicine and welfare economics which are theoretically underpinned by positivist and normative methods, respectively. More specifically, economic evaluations in the field of PA and SB, require skills and knowledge from a wide range of relatively new research fields which are underpinned by methods which aim to be practical (pragmatic). The pragmatic-orientated methods and topic areas which feature in the reflections in this thesis include: cost-utility analysis, cost-consequence analysis, trial methodology, estimation statistics, public health, behavioural science, PA and SB.

6.4.1.2. Assigning unit costs

A major theme in this thesis was that some of the methodological approaches outlined in the initial framework were not straightforward to apply in practice. This was despite my initial framework being informed by the methods used in existing studies identified in my systematic review (Chapter 2), searches carried out across the DIRUM database and the use of handbooks and best practice guidelines for the generic field of health economics. In particular, I found it challenging to standardise my costing methods within and across the Co-PARs and SLaMM trials. Different cost types required different sources of unit costs and additional calculations, and in many cases there was no standardised published data available so I was required to use shadow prices and make study-specific analyst-based assumptions.

6.4.2. Comparison of effectiveness and economic evaluations

6.4.2.1. Standardised and validated measurement tools

A key observation I made when trying to measure intervention, participant and productivity costs was that there was a lack of standardised and validated tools for complex public health interventions such as the Co-PARs and SLaMM interventions. Through this PhD project, I became increasingly aware that compared to the methods for effectiveness evaluations, economic evaluations measurement and reporting tools are underdeveloped. For example, in the field of economic evaluation, there is a lack of validated and standardised tools to measure resource use items for specific areas of health. By contrast, in the field of effectiveness evaluation, there is an initiative which specifically supports the standardisation and validation of health condition specific outcome measures. This project is called the core outcomes measures in effectiveness trials (COMET) initiative and aims to develop an agreed core set of outcomes for specific areas of health (COMET, 2020).

6.4.2.2. International vs country-specific guidelines

Another difference I observed was that analysts leading the effectiveness evaluations for the Co-PARs and SLaMM trial draw on international guidelines to inform the primary outcome measure and design of their evaluations. By contrast a number of key methodological steps I carried out for the economic evaluation of the Co-PARs and SLaMM trials were informed by country-specific guidelines (NICE, 2014a). In part, this is due to the different theoretical underpinnings of the two evaluation types. Prior to undertaking this PhD, I was unaware of these theoretical differences. For instance, clinical effectiveness evaluations appear to derive from the paradigm of evidence-based medicine which are traditionally underpinned by positivist methods. Meanwhile, economic evaluation stems from welfarism and extra-welfarism which are traditionally underpinned by the paradigms of normative methods, with extra-welfarism including pragmatic methods to determine how health states are measured and weighted. Nevertheless, the thesis highlights that economic evaluation studying complex lifestyle behaviours such as PA and SB incorporate draws of positivist, normative, pragmatic and naturalist methods, suggesting overall the field of study necessitates a pragmatic approach. Overall, my reflections on this topic helped me to recognise that it would be challenging for me to produce a practical framework that addressed all the steps involved in an economic evaluation and had international applicability.

6.4.2.3. Terminology

Another observation I made was that there is stark differences between the terminology involved in the conduct of trial-based economic evaluations and trial-based effectiveness evaluations. Having never conducted a clinical trial or an economic evaluation prior to commencing this PhD, many of these differences only became apparent in the field when I was applying my framework to the Co-PARs and SLaMM trials. For instance, the health economic literature includes terms such as perspective, opportunity costs, time horizon, discounting, cost-effectiveness planes, cost-effectiveness acceptability curves, incremental analysis.

6.4.2.4. Specialised training

During the data analysis phase of Co-PARS and SLaMM trials, I was surprised to find there was a dearth of practical guidance and examples in the economic evaluation literature on how to apply the principles of economic evaluations to the analysis of data from trials of complex PA and SB interventions. As a result, I completed a 3-day training course called 'Applied Methods of Cost-effectiveness Analysis in Health Care'. The course gave me 'hands on' computer-based experience on how to perform the statistical and uncertainty analysis for a trial-based economic evaluation. The practical methods taught on the course were reported as being standard practice in the UK. As the Co-PARS and SLaMM studies were the first trial-based economic evaluations I had conducted and as there is no published critical appraisal tools for assessing the quality of statistical and uncertainty methods in economic evaluations, I replicated the approach I was taught on the course. Furthermore, my supervisor who had expertise in health economics also helped me carry out and interpret the uncertainty analysis. This means I did not attempt to improve or change the approach I was taught by the course or advised by my supervisor. Nonetheless, I did reflect on the cost-effectiveness results including the uncertainty analyses such as the cost-effectiveness plane and cost-effectiveness acceptability curve. I recognised that interpretation of the results from these economic analyses differed to the interpretation of the results from clinical effectiveness evaluations. Overall, understanding the uncertainty in the results presented required specialist knowledge of health economic concepts which I feel would be challenging for non-health economists to understand without formal training or expert guidance.

6.4.3. Overview of actions to facilitate framework implementation

6.4.3.1. Documenting and reporting complexity

Ethnographic research in the form of observations and informal discussions can help researchers discover what the research is really about (Atkinson and Hammersley, 2007). I feel that the ethnographic notes and reflections I made during this PhD project enabled me to discover what contribution to practice my framework could make. It was through my practical experience of trying to design, implement and refine a novel standardised framework which covers the many complex steps involved in the conduct of an economic evaluation, that I discovered that even within the generic field of economic evaluations, some methodological steps are underdeveloped. I reflected on the scope of my PhD and recognised that I would only have the time, expertise and practical knowledge to make a novel contribution to a number of key methodological steps. Many of my reflections refer to the identification and measurement of resource use, this revealed to me that this is where my prior knowledge in public health can make an important contribution to the design of a

framework for assessing PA and SB interventions. My reflections also show how anthropological methods such as ethnographic data can support the analyst to document and explicitly report what was done and why. This transparency in methods and rationale is important since it helps the user of the results understand the context of the results, intervention and overall economic evaluation. The reporting guidelines published over the last decade indicate there is a drive towards transparency in trial-based research.

6.4.3.2. Describing, identifying and measuring resource use for complex PA and SB interventions

In my reflections, I refer to existing guidelines from the public health literature (e.g. TIDieR and PRECIS-2 frameworks) which are designed to capture complexity. Through this PhD I demonstrate how existing public health frameworks can support the practical steps in an economic evaluation for PA or SB. For example, the TIDieR framework is designed to improve the reporting of intervention descriptions. In Chapter 4 and 5, I found the TIDieR framework helpful for identifying intervention costs in terms of the types and quantify of resource use items. I also found the TIDieR framework was a more comprehensive way to describe a complex individual PA and SB intervention, when compared to existing health economic approaches, such as the pathway analysis approach. Pathway analysis involves specification of the intervention by asking the following questions: "who does what, to whom, when, where, and how often?" (Vos et al., 2007). Content of the TIDieR framework and pathway analysis approach are similar which indicates that from both a public health and health economics perspective, capturing detail and complexity about how the intervention is delivered is important. I found the TIDieR framework captures more detail as it includes items which require the analyst to distinguish between planned and actual resource use. Furthermore, my prior knowledge in anthropology has helped me recognise that the 'thick' description approach used in ethnographic research is also similar to the TIDieR framework and pathway analysis approach as anthropology's 'thick' description approach aims to 'provides clues to decide when, where, with whom, how and on which issues to intervene' (Krumeich et al. 2001: 216).

6.4.3.3. Applying health economic concepts to public health trials

Through my PhD journey, I have identified existing practices and concepts from the health economic literature which I believe can improve the quality of PA and SB trials. These include explicitly stating who the results are for, a concept known in health economics as stating the study perspective and presenting this in a table (as I did for the cost-consequence balance sheet). Furthermore, defining the perspective (where perspective is synonymous for stakeholder) in terms of what resource use and effects are likely to fall under the definition of each perspective, acknowledging that some categories such as productivity may be relevant to more than one perspective. As well as stating the study perspectives, another health economic concept which I believe can improve public health research is the concept 'opportunity cost'. In the Co-PARs and SLaMM trial, through informal discussions and meetings with the leisure centre and workplace staff, I became aware of which resources potentially represented an opportunity cost and were therefore a priority item to capture data on. For example, where the staff felt some or all of the time they allocated to deliver some of the intervention activities could have be used to deliver more important non-intervention related activity, then represent an opportunity cost from their

perspective. Furthermore, I believe my focus on the economic features of the trails helped me recognise that the planned interventions for the Co-PARs and SLaMM trials as reported in the protocols, had not captured all important resources (e.g. staff time, travel and private rooms) which were required to support the implementation of the intervention. I feel a health economic perspective may therefore help translate research findings into practice and address the 'implementation gap'.

6.5. Critical discussion of key themes

6.5.1. Standardising the first-step in the costing process

For more than a decade there has been a call for more methodological research in order to standardise costing methods (Mogyorosy and Smith, 2005). While this issue is too substantial for this thesis to address through a single framework, the previous chapters in this thesis have illustrated where standardisation is achievable. Standardisation is possible if resource consumption is quantified and presented in its natural units as seen in the cost-consequence balance sheets in study 2 and 3. This is because costing is a two-step process, involving firstly the quantification of resource use and secondly the assignment of unit costs. The reflections on piloting the framework in study 2 and 3 have illustrated that the second stage (assigning unit costs) of costing is the most challenging. Practically this stage involves numerous calculations, assumptions and secondary sources of varying quality as demonstrated in Appendices C.2 and D.3. Furthermore, as the authors of the studies reviewed in Chapter 2 do not discuss the calculations and assumptions they have made when assigning their chosen unit costs, it is likely that this stage of costing is analyst-dependent. Researchers and decision-makers unfamiliar with economic evaluation are unlikely to be aware of these challenges and thus without training in health economics, may not feel competent enough to critically appraise the costing methods performed. A consequence of this is a lack of transparency in how economic evaluations are performed which contradicts the standardised approach that the framework aims to achieve. Additionally, standardisation for the methods used to assign unit costs is complicated further by the fact that different countries and organisations have different healthcare systems and accounting practices (Cylus et al., 2016). Earlier efforts to standardise costing methodology for health research have not been successful (Busse et al., 2008).

Critically, costing methodology has not received the same attention as outcome methodology. Health economists have pointed out that when resource use data is collected from RCTs, it is often done inconsistently on a case-by-case basis and as a result there are calls for a research agenda, which focuses on improving methods for resource use measurement (Thorn et al., 2013a). Initiatives aiming to improve methods for measuring resource use, such as the PECUNIA and ISRUM projects, have been launched within the timeframe of this PhD, highlighting the timeliness of the findings in this thesis (PECUNIA, 2019, Thorn et al., 2018). The results of my systematic review (Chapter 2) as well as the process in developing the initial framework (Chapter 3) illustrated that resource use is being measured inconsistently both within and between countries, in terms of what resource use categories and items are deemed relevant and important to measure for various perspectives (multiple stakeholders). It has been argued that although the number of trials

collecting economic data has increased over time, the credibility and usefulness of this data is restricted by the variation in cost-effectiveness methods and reporting (Ramsey et al., 2015).

The revised version of my framework focusses on the methods for collecting and presenting resource (economic) data in its natural units alongside effectiveness data. By reporting disaggregated resource use and quality of life data, this may help make results more useful for other researchers, which will increase the visibility and impact of economic evidence. Systematic (and pragmatic) reviews can draw on this economic data, which can be used as input parameters for decision-analytic models (Sculpher, 2015, Briggs et al., 2006). Disaggregated data is typically input into decision-analytical models rather than meta-analyses of cost-effectiveness results. Anderson (2010) argues that the pooling of cost-effectiveness results is futile, due to the contextual differences between economic evaluations. Another advantage of presenting a breakdown of the resource use data is that reviewers can use this data to identify trends in terms of if, and how, outcomes vary by level of resource (Anderson, 2010). This information can also be used to inform the design of interventions in term of informing what type and quantity of resources optimise the outcomes of interest. Abu-Omar et al. (2017) suggests a similar analysis, as they argue that PA studies should assess whether the type of staff delivering an intervention can shape whether an intervention is cost-effective. Bryan and Williams (2014) propose this type of economic research as being knowledge-generating. They argue that a more positive (as opposed to normative) approach to the evaluation of economic data, could generate results that are meaningful to a broader range of decision-makers including researchers designing interventions. Others have recognised that there is scope for the discipline of economics to contribute to the design of interventions in order to bring about change at the macro- and micro-level (Frew et al., 2018).

6.5.2. Presenting resource use data alongside effectiveness data

Research has shown that when economic evaluations are simultaneously published with results from the effectiveness evaluation, this is likely to improve overall dissemination of the economic evidence (Thorn et al., 2013b). Although, publication bias is a known phenomenon for effectiveness results (Song et al., 2000), publication bias and delays in publication are even greater in the field of economic evaluations (Greenberg et al., 2004). The review by Thorn et al.(2013b) observed a two-year delay in publication when economic evaluations are not published alongside the effectiveness results. The reviewers claim this as concerning, arguing that a two-year window is enough time for effectiveness results to bring about change in practice and reimbursement decisions without consideration of the economic evidence. A recent study of a workplace SB trial published in a high impact journal provides a good example of how the economic-related variables quality of life, absenteeism and presenteeism can be quantified and presented in the same publication as the effectiveness results (Edwardson et al., 2018). More specifically, the study briefly describes the main results in the main manuscript and signposts the reader to the supplementary materials for a transparent breakdown for each comparator group. That said, the authors state the cost-effectiveness results will be reported in a future study, which may explain why the resources required to set up and deliver the intervention are not comprehensively quantified. If future systematic reviews therefore want to explore how levels of resource use relate to levels of outcome, the reviewers will be required to contact the authors, which may be challenging if this is

some time after the study. Alternatively, they would have to make assumptions about the levels of resource use, which will be less precise prospective study records. For almost two decades it has been argued that a range of costs and consequences should be presented in trials, including quality of life and intervention costs as these are relevant to decision-makers and practitioners (Tunis et al., 2003). The CCA makes it explicit what costs and consequences have been included and omitted, which overall means the results are more likely to influence real-world decision making as decision-makers can apply their own values to the results so as they align better to their local context (Coast, 2004).

6.5.3. Applying existing public health methods to the field of economic evaluation

One way to reduce the researcher burden from collecting resource use data could be to draw on tools that are already used when assessing PA and SB intervention. This thesis has shown, that the TIDieR framework is useful for collecting data on intervention staff type, materials and capital equipment needed to carry out an intervention and that the PRECIS-2 framework is useful for understanding and quantifying the resources associated with the contextual factors. At present, analysts are only required to report this qualitatively when they submit their journal papers. There is scope to quantify this data further by using the microcosting tool, which has been developed and refined during the evaluation of the Co-PARS and SLaMM trial. The refined version of the microcosting tool and examples of its application are provided in appendix E.1 and E.2.

6.5.4. Complexity in assigning costs to resources

'Collecting cost data in a manner that is simultaneously concise, understandable for patients, valid, precise, consistent between trials, and generalizable is challenging.'

(Thorn et al. 2017,p648)

The quote by Thorn and colleagues refers to the present challenges even researchers trained in health economics face when collecting economic data alongside trials. Fundamentally, this challenge is likely to be due to the fact costing methodology is deeply underdeveloped. In the SLaMM and Co-PARS trials, participants were asked to self-report the secondary care departments they were admitted to and visited for their outpatient and day case appointments. Similarly, participants were asked to name the type and dose of the medication they were being prescribed. Data were collected as it was expected that it would be a straightforward exercise assigning unit costs to the secondary care and medication received. The UK has comprehensive and standardised databases, which list the unit costs for hundreds of different treatment and medication types offered within the NHS (NHS Improvement, 2018, NHS Digital, 2018). The selection of an appropriate unit cost was time-consuming and relied heavily on the researcher's (Madeleine Cochrane) judgement. The idea that cost-effectiveness studies may not be cost-effective themselves is discussed by Drummond et al. (2015b) who claim that costing is a skill which involves a trade-off between precision/accuracy, and research effort/ feasibility. That is to say, this may have been the case when conducting the Co-PARS and SLaMM trials, as the

calculation of individual unit costs was a time-consuming process. Furthermore, feasibility in relation to costing methods, has been described as being the 'ability to observe' the resource use as well as the cost of collecting it (Mogyorosy and Smith, 2005). Others have commented on this, claiming that precision can be compromised by the availability of data and time for the study (Kinsella, 2008). This can be understood by considering whether there is 'enough precision for the decision' (Lipus, 2018). It is argued that less accurate resource use estimates may suffice for a particular decision (Mogyorosy and Smith, 2005). Thorn et al. (2018) draw attention to this challenge faced by researchers: 'there is a trade-off between gathering as much information as possible (with increased patient burden and possible poor response rates) and gathering less information (which may not allow an accurate analysis to be conducted)'(p641) (Thorn et al., 2018).

6.5.5. Timely and relevant data for public health stakeholders

As discussed in my systematic review (Chapter 2), one of the challenges decision-makers face when investing in preventative programmes is attributing long-term costs and consequences to the intervention under study. In general, there is a lack of longitudinal evidence identified to accurately estimate the long-term impacts of PA and SB interventions. However, the Co-PARS trial indicated that an increase in quality of life and reduction in GP consultations may be observable over a shorter timeframe, in this case the 6-month trial time horizon. The ability of the framework to consider the immediate resource impacts (costs) may increase the role of economic evidence in public health commissioning. A study exploring the perspectives of the Directors of Public Health across local authorities in the UK, found a need for more evidence on the short-term economic impact of the public health interventions they commission, in order to make the case for investing in preventative interventions to local government (Willmott et al., 2016). NICE's most recent methodological guidelines recommend that economic evidence on short-term evidence from trials should be included in economic evaluations (NICE, 2014a). Importantly, the short-term data is likely to be relevant to public sector decision makers who operate on a short-term budgetary cycle (PHE, 2018, Willmott et al., 2016). If more immediate resource use and quality of life evidence is generated in the short-term, there is a need to ensure this data is disseminated amongst those who can use it.

6.5.6. Evidence fit for a non-health economic audience

The summary cost-effectiveness and uncertainty results of the primary analysis (CUA) in the SLaMM and Co-PARS trials presented complex information that was not easy to understand without training in health economics. By contrast, the disaggregated results and cost consequence balance sheets are easier to interpret for each stakeholder (perspective). In addition, this approach promotes transparency. As seen in the Co-PARS and SLaMM trial, it was clear to see what costs and consequences each stakeholder experienced. This balance sheet complements the recent drive towards systems thinking within the public health (Rutter et al., 2017) whereby the perspective of multiple sectors are taken into account. Evidence that is easy to interpret is important to both the researcher and decision-maker if resource allocation and intervention design are to be informed by evidence. There is a need to ensure that the results generated by studies which use the refined framework are relevant and accessible to the reader. A consequence of not presenting evidence that is clear and easy to understand is that the decision-makers cannot be

fully accountable for their decisions (Bryan and Williams, 2014). Owen et al. (2017) argue that it is important that decision-makers understand why cost-effectiveness results vary across studies. Some decision-makers expect CUA and CEA studies to be performed as 'add on' exercises alongside effectiveness evaluations. This places an unrealistic expectation on researchers, since this thesis has shown that costing methodology is underdeveloped and consequently involves many analyst-based decisions.

Evidence suggests that local level decision-makers firstly, do not always understand the health economic analyses they are presented with and secondly, find the recommendations from cost-effectiveness studies are unrealistic for 'real world' decision-making (Bryan and Williams, 2014). For instance, some have difficulty in redirecting resource use, while others are under pressure to contain costs due to the nature of the short-term budget cycles they operate within (Bryan and Williams, 2014). A recent Delphi study aiming to capture public health decision-makers views on economic evaluation found that the decision-makers reported that cost-effectiveness ratios were not helpful, they preferred the results to be presented transparently in terms of costs and outcomes being presented by sector and population group (Frew and Breheny, 2019). A previous study also claimed QALYs are not well understood by decision-makers (Drummond, 2003). Cylus et al. (2017) describes cost-effectiveness results which are reported as "resource use per unit of health system output" as being "beguilingly simple" as it does not account for the complexity of the health system and is only looking at one independent variable in isolation.

6.6. Implications for research practice

6.6.1. Capturing complexity through informal conversations with stakeholders

Recent discussions around the methodological challenges experienced in the wider economic evaluation literature indicate that the identification of costs and consequences from multiple sectors, as discussed as one of the four key challenges in my systematic review (Chapter 2), may no longer be regarded as being a challenge unique to public health trials (Ramsey et al., 2015, Drummond et al., 2015b). This suggests the discussions within this thesis may be informative to researchers and decision-makers in the wider field of economic evaluation. During the design stage of future economic evaluations, analysts could conduct multi-stakeholder and expert consultations to map out which costs and consequences (Squires et al., 2016) are relevant to the CCA. A systems thinking approach (Rutter et al., 2017, Squires et al., 2016) is recommended to ensure interventions' indirect and unintended costs and consequences on multiple sectors are considered, not just those experienced by the health sector or payer. Two recently published frameworks can help analysts apply a systems approach (Cylus et al., 2016, Squires et al., 2016).

6.6.2. Adopting existing tools from public health

A surprising finding which emerged in my systematic review (Chapter 2), was that authors of existing PA and SB economic evaluations did not reference or provide templates in the supplementary materials for the type of measurement tool they were using to measure resource use. More specifically just one (Edwards et al., 2013b) of the 15 studies reviewed in Chapter 2 provided a reference to the tool they used to measure healthcare utilisation. Together this evidence shows there is potential for researchers from multiple disciplines to develop and pilot resource use

measurement tools in their future PA and SB trials. This can be developed and shared free of charge via the DIRUM database. A key learning from this thesis, was that there are existing multidisciplinary standardised tools within the clinical effectiveness literature such as the TIDieR and PRECIS-2 tools which implicitly incorporate concepts that are relevant to economic evaluations. In particular, these tools may be helpful for capturing intervention costs and contextual factors respectively. In order to reduce the burden on trialists collecting resource use quantities throughout their trial in terms of healthcare utilisation, Chapter 4 and 5 identified resources that are more likely to be cost drivers. For primary healthcare the Co-PARS and SLaMM trials identified the same priority resource items, this included the GP, Practice Nurse and Counsellor. In addition, visits to the Physiotherapist were also identified as being relevant to the Co-PARS trial. That is to say in future studies, if time and resources are limited, researchers could prioritise data collection for these cost items.

All five models in Chapter 2's review presented a visual depiction of the disease pathway for PA. Researchers could draw on the disease pathways presented in the model-based studies in Chapter 2 in order to help policymakers and those designing interventions populate a logic model with the long-term costs and consequences associated with PA and SB. The visual presentation of a logic model is recommended in public health effectiveness evaluations (Moore et al., 2015). More systematic evaluation of care pathways is advocated to find efficient ways to integrate behaviour change and treatment services, to prevent major chronic diseases such as T2D (ISPOR, 2018). A visual developed during the design stage of the trial can make it clear to the researchers and commissioners of the evaluation how immediate investments to PA interventions have the potential to make substantial savings in the future, through the aversion of costly treatment for CHD, T2D and stroke (Campbell et al., 2015b). To date, even when cost-effectiveness has been incorporated in a logic model of a high quality study, this has typically been limited to stating 'intervention is cost-effective' rather than specifying what costs may be modified (Edwardson et al., 2018).

6.6.3. Supporting practice for early career health economists

It was originally anticipated that all researchers from the wide range of disciplines in which PA and SB interventions include, as well as health economists, would be able to use the framework for their upcoming PA and SB trials. However, as numerous examples across the thesis have brought to light, the methods and terminology involved in economic evaluations differ substantially to effectiveness evaluations. As a researcher who had no formal training in economic evaluation before commencing this PhD, I found there was a lack of detailed practical guidance in the literature to guide me through the practical steps in conducting an economic evaluation. This meant I was unable to implement all steps in my planned framework without receiving additional formal training in economic evaluation. I was access a 3-day practical course on economic evaluation principles applied in practice. It would have been inappropriate for me to repeat the content of the training course in my framework especially as my approach would not have been novel and furthermore I believe the most important aspects of the course were the practical computer-based exercises which went beyond the scope of my framework. Through the development and implementation of my framework I discovered my framework has the potential to make an important contribution to early career health economists who do not have prior knowledge in public

health, but are required to conduct economic evaluations of individual-level PA and SB interventions. More specifically, through my reflections I discovered that existing public health guidelines (e.g. PRECIS-2 tool, TIDieR checklist) and literature are helpful for the economic evaluation steps that involves the identification of relevant perspectives and resources. I have also discovered that anthropological methods such as observations and informal conversations with the various stakeholders involved in the commissioning, set up, delivery and hosting of the intervention are helpful when measuring the amount of resources used or saved.

6.6.4. Standard Operating Procedure

In order for an organisation to be registered as a clinical trials unit by the UK Clinical Research Collaboration (UKCRC), the organisation needs to demonstrate they have a number of key competencies. The UKCRC recommend the use of standard operating procedures (SOP) to demonstrate the key competency of quality assurance (UKCRC, 2019). SOPs are documents, which explicitly specify an organisation's preferred approach to carrying out a specific trial procedure (e.g. randomisation, database, document control, data management, statistics). The aim of SOPs are to achieve consistency, efficiency and quality in a specific procedure across an organisation. The UKCRC lists 19 procedures for which they argue it is essential to have a SOP for. Trial units across the UK usually have their own SOPs for the various procedures. They also play a key role in supporting communication between multidisciplinary teams. At present, health economic SOPs are not essential, however some argue that health economic SOPs would support the integration of health economic procedures into clinical trials (Edwards et al. 2008). The intention of my framework presented in chapter 3 was to provide a multidisciplinary and pragmatic framework for economic evaluations in the field of PA and SB. Through my PhD journey I have become aware of the differences in methods and terminology of effectiveness and economic evaluations. I believe an interface which supports multidisciplinary working in public health economics, such as a SOP, is required to support the communication and implementation of my framework for identifying and measuring resource use. The SOP I present in the subsequent section draws on my reflections on developing the initial framework (Chapter 3) and applying the framework to the Co-PARs (Chapter 4) and SLaMM trial (Chapter 5).

6.7. Presentation of the standard operating procedure

Standard operating procedure for resource use measurement in the economic evaluation of individual-level PA and SB interventions

1. Purpose

The purpose of this standard operating procedure (SOP) is to describe a non-study specific approach to the planning, collecting and presenting of resource use data for individual-level PA and SB trials. The aim of this SOP is to standardise some of the routine steps in the conduct of an

economic evaluation of individual-level PA and SB interventions. The SOP describes a minimum set of perspectives, resource categories and equity subgroups for the field of PA/SB.

2. Scope

This SOP can be applied to funding applications, protocols, ethical approval processes, health economic analysis plans, interim and final reports, and journal publications. The SOP is intended to: (1) support early career health economists applying economic evaluation to the field of PA and SB; and (2) act as an interface between health economic analysts and a multidisciplinary trial team. The SOP is complementary to country-specific guidelines for economic evaluations. It does not replace country-specific guidelines since these are generic guidelines to the field of health. Instead the SOP provides guidance for analysts to carry out additional analyses which could be specific to the PA and SB community.

3. Introduction

Economic data in terms of resource use and generic quality of life data (final health outcomes) is important for ensuring society's scarce resources are being allocated to interventions which achieve the greatest value for money. The complex design of individual-level PA and SB trials demands health economists have an understanding of at least the following culture and context specific factors: (1) the organisational structure of the setting in which the PA and/or SB intervention is set (this is likely to be a non-healthcare setting); (2) the expertise, materials or capital (e.g. private rooms) at the setting and whether these resources represent an opportunity cost; (3) the various perspectives (stakeholders) who may be affected intentionally and unintentionally by the PA and SB intervention; and (4) the immediate and future costs and health benefits associated with individual-level PA and SB interventions.

Health economists new to carrying out economic evaluations in the field of public health can draw on the range of pre-existing peer-reviewed public health tools which are designed to support researchers in the conduct of trials which are fit for purpose (pragmatic). Similarly, health economists may benefit from methods used in anthropology. For example, by providing 'thick' descriptions of the context in which an intervention is delivered this may help improve the analyst's understanding of the resource implications of implementing and delivering a PA and/or SB intervention. Furthermore, health economists could make use of the information they receive about the intervention and setting context from the informal conversations they have with stakeholders and multidisciplinary experts during the study period.

The different terminology and methods used in economic evaluations compared to effectiveness evaluations demands trialists to have an understanding on what and how economic data is important to measure. Trialists and health economists should require a trial-based tool, such as a SOP, to ensure an economic perspective is incorporated into the trial design, to monitor data completeness and to support dissemination of evidence on resource use. Section 6 of this SOP outlines a procedure to improve the conduct and communication of an economic evaluation of a individual-level PA and/or SB trial. The procedure provides detailed advice on how some key steps in economic evaluation could be carried out in practice. The SOP draws on insight and existing

practical methods from the disciplines of public health, anthropology and health economics. Practical templates and examples to support implementation of the SOP are provided in Appendices E.1-E.3.

4. Definitions

4.1. Definitions of key public health terms

Stakeholders: individuals or organisations who are likely to be affected by the intervention of interest. Identifying stakeholders is similar to stating the study perspective.

Logic model: a conceptual model which graphically depicts and hypothesises the relationships between inputs, activities, outputs, and short-, medium- and long-term outcomes of typically one intervention of interest. Typically, public health logic models report improved health and wellbeing as the final outcomes. Long-term cost implications, length of life and quality of life are not typically incorporated in public health logic models, but they could be.

Equity: A key objective of public health interventions. Equity refers to the unfair distribution of health and income in society.

4.2. Definitions of key anthropological terms

Ethnographic data: sources of data typically collected in an informal and unstructured manner from everyday interactions including day to day observations and informal conversations. The aim of ethnographic data is to explicitly document how complex the ‘real world’ is rather than simplify it.

Ethnographic reflections: descriptions of what happened (i.e. what actions were taken) and an interpretation on why this happened (i.e. why these actions were taken). Interpretative content is informed by the detailed descriptions and observations recorded from everyday interactions (the ethnographic data). This documentation is similar to the documentation that is increasingly requested by journal reporting guidelines. For instance, many reporting guidelines necessitate that analysts transparently report what exact methods they chose and why.

4.3. Definitions of key health economic terms

Economic evaluation: comparison of both costs and effects of at least two intervention (or policy) options

Costing: a two-step process. The first-step involves identifying the type and amount of resources used. The second section involves assigning an appropriate unit cost. Published national unit cost data sources are recommended in the UK, although most published data is for healthcare which makes assigning costs to interventions delivered in non-healthcare settings particularly challenging.

Opportunity cost: the benefits given up as a result of choosing one course of action (e.g. an intervention activity) over another. For example, when resources (e.g. time, materials, capital) are used up to support the operation of one intervention, then these resources are no longer available at that point in time for another intervention. The potential benefits of the intervention which is not

delivered represents the benefits that are foregone. In health economics the foregone benefits are called the ‘opportunity cost’.

Time horizon: time period in which costs and outcomes are analysed over. In a trial-based economic evaluation, the time horizon is typically the study follow up time (e.g. 12 months follow up).

Perspective: the individuals or organisations who are likely to be affected by the intervention of interest. Stating the study perspective is similar to the identification of relevant stakeholders.

Incremental analyses: analysis of the between group difference in the outcome (costs) of interest. Some health economists may use the term ‘additional benefits (costs)’ to describe this.

Incremental cost-effectiveness ratio (ICER): the most commonly reported summary statistic for economic evaluations. The ratio represents the additional difference in costs of two intervention options divided by the additional difference in effects of the options.

Willingness-to-pay (WTP) threshold/ Cost-effectiveness threshold: The maximum amount a decision-maker is willing to pay per unit of benefit.

Decision-analytic model: an analytical methodology that draws on primary and secondary data sources to estimate the costs and effects of at least two courses of action (e.g. intervention or policy strategy) for a pre-specified population group over a pre-specified time period. Model parameters are similar to the variables included in logic models but also include cost implications, length of life and quality of life.

5. Responsibilities

5.1. Health Economist

It is the responsibility of health economist to implement the procedures outlined in the SOP. In particular, it is the responsibility of the health economist to draw on the data from this SOP and specify in the trial documents (e.g. protocol and health economic analysis plan) what resource use data is relevant and important to collect for each PA/ SB trial. The final decision on how resource use data will be collected and quantified will be decided by the health economist.

5.2. Trial team

It is the responsibility of the trial manager and trial team involved in the collection and monitoring of the trial data, to read through this SOP so as they understand the overarching approach that will be applied for the generation of resource use data. It is expected that the trial manager and trial team will provide advice about the practicability of the data collection methods proposed by the health economist. In addition, the trial team can support the health economist to document the complex contextual factors associated with the intervention and trial. This information can be recorded and discussed with the health economist through informal ongoing trial meetings.

6. Procedure

The following 13 items listed in Table 1 in section 6.1 and the guidance provided in section 6.2 are intended to support a multidisciplinary and pragmatic approach to the identification and measurement of resource use associated with individual-level PA and/or SB trials.

6.1. Overview of key steps to address

Table 1. Overview of items included in the revised framework

Items*	
1	How can complexity be incorporated into the studies generating evidence on resource use?
2	How can studies generating evidence on resource use define their study question?
3	How can important and relevant resource use be identified?
4	How can intervention set up and operating costs be measured?
5	How can health and social care resource use be measured?
6	How can participant-related resource use be measured?
7	How can employer-related resource use be measured?
8	How can resources be valued?
9	How can summary data can be presented?
10	How can equity be incorporated into the analysis?
11	How can the study results be interpreted?
12	How can resource use evidence be disseminated?
13	How can the resource use evidence be used to support public health practice and research?

* Adapted from the methodological quality assessment checklist outlined in the Drummond et al. (2015) and informed by the empirical studies (Study 1-3) from the thesis by Cochrane, 2020.

6.2. Explanatory guidance to support implementation of the SOP

Item 1. How can complexity be incorporated into the studies generating evidence on resource use?

As economic evaluations require data from studies which compare at least two comparator groups, clinical effectiveness trials represent a key opportunity to collect primary data on resource use. It is advisable that a researcher with expertise in health economics (e.g. health economist) is involved in the trial from its inception. This is to ensure a health economic perspective and economic data collection is incorporated into the study design. Key activities at the start of the trail, may include a health economist participating in both the formal trial management group meetings as well as the more informal conversations around the trial planning and set up. Together the health economist and trial team could commit to the collection and presentation of resource use data by stating it in the trial's registered protocol.

PA/ SB interventions are typically set in non-healthcare settings. It is advised that the health economist captures the context in which the trial is being implemented, this could be captured through ethnographic data and reflections (see definitions in section 4.2 of this SOP). In order, to

gather contextual data in a systematic way, it may be helpful to reflect on the nine domains from a tool called the Pragmatic Explanatory Indicator Summary 2 (PRECIS-2) tool. PRECIS-2 derives from the trial methodology literature. It encourages researchers to explain their trial design decisions and to reflect on whether their trial is fit for purpose. The nine domains could be used to structure any informal discussions with the trial team around context and the complexity of the intervention and trial setting. Domains from the PRECIS-2 tool which are particularly relevant from a health economic perspective include: (1) the setting- where is the trial being delivered?; (2) the organisation- what skills, expertise and other resources are needed to deliver the intervention?; (3) the outcomes- how relevant are the outcome measures to the individual participants and other stakeholders?; and (4) the analysis- for instance, for economic evaluations it is recommended that the intention-to-treat (ITT) principle is followed, whereby all participants are analysed according to the group they were assigned to, regardless of whether they have adhered to their group's protocol.

Item 2. How can studies generating evidence on resource use define their study question?

Defining the study question is similar to an approach used in effectiveness evaluations where the population, intervention, outcome, comparator, study type (PICOS) framework is commonly used to define the components of a well-defined study aim. In health economics, the following five pieces of information are recommended when defining the study question: study type, study technique (how results are presented), study perspective, time horizon and population (Drummond et al. 2015). An applied approach for addressing these five factors in practice is outlined below:

1. *Study type:* It is recommended that the team explicitly state in the trial protocol that both resource use and effects are being collected from at least two comparator groups. Comparator groups may include a new or refined PA/SB intervention, no intervention, or a group that represents usual care for the study setting.
2. *Technique (summary results presented):* It is recommended that resource use categories and outcomes are analysed and presented separately in a disaggregated format. In addition, present resource use and outcomes according to the stakeholder (perspective) in which they related to. This is similar to the way a cost-consequence analysis is typically presented.
3. *Study perspectives:* It is helpful to define the study perspective by making a list of all the individuals and organisations you think may be impacted by the implementation of a PA/SB intervention (see item 3 of the SOP for more detailed guidance on this step).
4. *Time horizon:* It is helpful if the data collection and analysis of the trial-derived resource use data is done over the same time schedule as the effectiveness data. A six-month recall period is deemed acceptable for participants taking part in PA/SB trials.
5. *Target population and subgroups:*. For practical and efficiency reasons related to the data collection schedule, it is recommended that the target population for the economic data is the participants recruited for the clinical effectiveness trial. One way to incorporate equity into the study could be to do subgroup analyses for the following subgroups: age, sex, socioeconomic status and pre-existing medical condition.

Item 3. How can important and relevant resource use be identified?

It may be helpful for the trial team to do this step together in order to ensure all important and relevant resources are considered. This could involve two key steps. Firstly, it may be helpful to stimulate discussion by referring to Rutter et al.'s (2018) complex systems thinking framework for PA. In part, the framework has been designed to help researchers understand the multiple sectors and stakeholders who may be related to the challenge of physical inactivity. The complex systems framework for PA is freely available in the public health literature. Secondly, as a team it may be helpful to identify and list all the individuals and organisations who are involved in: paying, providing, hosting, participating and setting up the PA/SB intervention. Individuals and organisations may fall into more than one of these categories (e.g. if the intervention is set in the workplace, the employer represents the host organisation and employer's perspective). The following types of perspectives and their associated resource use could be discussed:

1. The payer's perspective may include the planned intervention costs such as the resources required to set up and deliver the PA/SB intervention as specified in the protocol
2. The provider's perspective may include the resources used to set up and deliver the intervention which are not captured or differ to those specified in the protocol (e.g. additional preparatory time, ongoing support time)
3. The host organisation's perspective (the setting's perspective) may include the resources used to set up and deliver the intervention which are not captured or differ to those specified in the protocol (e.g. the use of private rooms to deliver the intervention).
4. Health and social care costs may include primary healthcare activity (e.g. consultations with the GP, practice nurse and allied health professionals). In addition, it may include secondary healthcare activity (e.g. A&E, inpatient, outpatient and day case visits to the hospital).
5. The participant's perspective may include the participant's out of pocket expenses (e.g. expenses for clothing, travel, gym membership) and productivity losses (e.g. loss of leisure time, loss of earnings) which accrue due to taking part in the PA/SB intervention.
6. The employer's perspective may include the indirect losses and/or gains in absenteeism and presenteeism. If the intervention is set in the workplace, time given up (and potentially income loss) due to staff participating in the intervention during working hours may also be captured.

Item 4. How can intervention set up and operating costs be measured?

As individual-level PA and SB interventions may represent a new intervention model which has not been evaluated from a health economic perspective, it is recommended that a microcosting exercise is carried out. That said, this may depend on the time and purpose of the economic evaluation. At a minimum, it is advised that the health economist check their interpretation of the interventions of interest with the trial team. Microcosting is recommended as it is expected to be more accurate that aggregate costs or block contract costs. Microcosting has a role in identifying

inefficiencies in resource use as well as capturing additional resource use that may not have been previously captured. The latter is important, as this 'hidden' resource use may help explain the phenomenon known as the 'implementation gap'.

Microcosting involves describing in detail the interventions of interest. In addition to describing the interventions of interest, it is recommended that the same description is provided for the comparison groups, even if these represent usual care. This is because usual care may vary geographically and/or by organisation. It may be helpful to base the intervention and usual care descriptions on items from the Template for Intervention Description and Replication (TIDieR). TIDieR is a reporting tool from the public health and trial methodology literature which encourages researchers to describe: what activities and equipment are required for an intervention; who delivers the intervention and how; where is the intervention delivered; is the intervention tailored or modified; and is the intervention delivered as planned (Hoffmann et al. 2014). An adapted version of the TIDieR framework is provided in the supplementary material (Appendix E.1). The adapted version provides further questions in order to help the research to quantify the resource use data. Furthermore, the tool categorises resource use into three broad categories: people-related (e.g. expertise and time), materials (e.g. resources specific to the intervention), place-related (e.g. existing resources at the host organisation not specific to the intervention, travel to the location).

Intervention resource use could be captured through a range of data sources including ethnographic data such as informal conversations and observations (see definitions in 4.2), diaries or logs, as well as more formal approaches such as face to face meeting at the intervention setting with the trial and intervention staff. If possible, collect this data prospectively. Prospective data collection is preferred as it is expected to be more accurate since it does not rely on human recall. Nevertheless, it may be necessary to be flexible on how and when this data is collected as individuals and organisations providing this data may have a preferred option especially if they perceive the collection of this type of data as being a burden. Resource use quantities gathered over the course of the study can be collated into a microcosting spreadsheet in a time stamped computer. An example on how this data could be presented is provided in Appendix E.2.

Item 5. How can health and social care resource use be measured?

It is advisable that this data is captured at the same baseline and follow up time points as the effectiveness evaluation. In the UK, methods for extracting resource use data from primary and secondary electronic medical records are expensive and underdeveloped. It is therefore advised that at a minimum, a self-report questionnaire is used to collect health and social care data. Questionnaires are inexpensive and can be easily incorporated into the questionnaire booklet for the effectiveness evaluation. The self-report questionnaire presented in Appendix B.1 is recommended as it has been successfully applied in practice to a PA and SB trial. The structure of the questionnaire is based a validated healthcare utilisation questionnaire called the Client Service Receipt Inventory (CSRI) (Beecham and Knapp, 1999). Adaptations to the original questionnaire include the addition of definitions for the secondary health care questions (e.g. defining inpatient

care) so as to support participants interpretation of the questions. A range of health and social care professionals are included in the primary care section of the questionnaire. Nonetheless, if there is a need to shorten the questionnaire (e.g. due to participant burden or limited resources), then the following types of primary care consultations can be prioritised as preliminary evidence (from Chapter 4 and 5) indicates these are the most commonly used primary care professionals amongst physically inactive and highly sedentary individuals: GP, Practice Nurse, Counsellor and Physiotherapist.

It is beyond the scope of this SOP to advise on how a decision-analytical model could be developed for PA/ SB interventions. If time and resources are limited, an alternative could be for the health economist to add resource use variables to a logic model. Logic models are increasingly cited in the public health literature. In brief they are designed to encourage researchers to consider how inputs, activities, outputs and outcomes. Outcomes are typically broken down in to short-, medium- and long-term outcomes. This degree of detail on resource use is not typically incorporated in logic models presented in the effectiveness evaluation. It is possible that this may be related to the observation that health economists are not usually involved in the development of a logic model.

Item 6. How can participant-related resource use be measured?

It is recommended that participant resource use data is captured through self-report methods and at the same data collection time points as the effectiveness evaluation. The participant cost questionnaire presented in Appendix B.1. may be helpful for capturing this data. This questionnaire is based on a pre-existing self-report questionnaire (Wordsworth and Thompson, 2001) retrieved from the Database of Instruments for Resource Use Measurement (DIRUM) and has been adapted to include examples of resource items related to PA/ SB interventions. The DIRUM database is a relatively new initiative by a group of trial-based health economists in the UK who aim to share measurement tools they have designed and applied to trials from various disease areas. At present, there is a lack of tools on DIRUM which have been applied to the context of individual-level PA and SB interventions. At minimum it is helpful to ask participants to estimate on average, the following:

1. time they spent taking part in and travelling to the PA/SB intervention, and whether this time represented an opportunity cost (i.e. did they give up leisure activities or incur a loss of earnings in order to take part which is deemed more valuable that the intervention)
2. whether their participation in the intervention led to a change in expenses (e.g. they purchased trainers, parking expenses, gym membership). As PA and SB are multi-dimensional behaviours it is also helpful to request that the participants report the units alongside the expenses they report. More specifically, units can be requested by asking the participants to report the following information (examples can also be provided to support the participants understanding): type of purchase (e.g. swimming session), the duration of purchase (e.g. 30 minutes of swimming) and the frequency (e.g. 5 times per month for three months). These questions align with the FITT framework from the PA and

SB literature which request PA and SB is described in terms of frequency, intensity, time (duration) and type (Barisic et al. 2011)

3. whether participation led to an increase or decrease in absenteeism and presenteeism which in turn led to an increase or decrease in income for the participant (i.e. this question may be most relevant for participants employed through casual and/or temporary contracts).

Item 7. How can employer-related resource use be measured?

If the intervention is not delivered in a workplace setting, the most efficient way to collect data on productivity from the employer's perspective is likely to be from the participant's self-report questionnaire. More specifically, the question which asks the participant whether they missed time in work due to participating in the intervention and whether this work time lost was a loss in earnings for the participant or employer.

If the intervention is set in the workplace, then gain or loss in productivity as a result of staff participating in the intervention can be captured through a range of methods. Firstly, productivity is a multi-dimensional activity including the two broad dimensions absenteeism and presenteeism. It is recommended that the health economist and trial team work together to decide on what productivity measures are most appropriate. There is a range of validated questionnaires available in the clinical effectiveness literature for capturing absenteeism and presenteeism. By contrast, there is a lack of consensus in the health economic literature on how to incorporate productivity into economic evaluations and a lack of guidance on how to collect objective absenteeism data. Therefore, in the absence of consensus and guidance, the most appropriate approach is likely to be one where a range of data sources are explored early on during the trial so as the best quality data source can be prioritised for the remainder of the study. Quality is likely to be determined by the completeness of the data. Early on during the trial it is advisable that the health economist and trialists see if it possible to access a sample extract of the company's electronic absenteeism records. Similarly, it is recommended that the completeness of the self-reported absenteeism and presenteeism questionnaires are also monitored early on.

Item 8. How can resources be valued?

This SOP focusses on the first step in the costing process which is the measurement and quantification of resource use (see example in Appendix E.2). It is helpful to present means and standard deviations for resource use data reported as continuous data (see example in Appendix E.3). If sample sizes are large enough it is advisable that the summary results are presented for the following equity subgroups: socioeconomic status, age, sex, medical condition.

Item 9. How can summary data can be presented?

Incremental differences between groups can be presented for each resource and outcome variable. In addition, it is recommended that means and standard deviations are reported for all continuous variables and proportions are reported along with the numerator and denominator for all

categorical variables. It may be helpful to display the summary resource use data in tabular form in accordance to the perspective they relate to. This table could resemble the format used for cost-consequence balance sheets (an example of the resource use table is provided in Appendix E.3).

Item 10. How can equity be incorporated into the analysis?

As equity is a key objective of public health trials, it is helpful to pre-specify which characteristics will be explored through subgroup analyses (also referred to as an equity subgroup analysis). It may be helpful to refer to the PROGRESS-Plus checklist from the public health literature (O'Neill et al. 2014). In particular, it may be helpful to focus on characteristics which have been most commonly assessed across other studies from the field of PA/SB, these include: age, sex, socioeconomic status and pre-existing medical condition. It is recommended that the health economist and trialist collect this data at baseline alongside the other demographic data (e.g. via the participant's questionnaire booklet). A key method used in public health practice for capturing socioeconomic status in the UK is by collecting data on postcode or place name. Postcode or place name data can be input into the Index of Multiple Deprivation (IMD) database (Department for Communities and Local Government, 2015) in order to interpret whether a person lives or works in an area of high deprivation. If it is deemed inappropriate to collect postcode or place names, alternative definitions for socioeconomic status could be used such as level of education, employment type.

Item 11. How can the study results be interpreted?

Observational notes and informal conversations (ethnographic data) gathered throughout the trial period can be reflected on during the trail. It is recommended that the health economist and trialists informally meet as a team to reflect on why methodological choices were made. These reflections can add contextual information to the study's discussion around the generalisability of the results. As discussed in item 1, the nine domains of the PRECIS-2 tool may help the trial team work through the trial's characteristics and reflect on whether they believe the findings are generalisability to other settings in the UK and transferable to other countries outside of the UK. For discussions on transferability it may be helpful to consideration of the funding and organisational structure of the multiple sectors involved in the PA/SB intervention. For example, in the UK health and social care is free at the point of use and publicly funded through tax.

Item 12. How can resource use evidence be disseminated?

It is advisable that the resource use data is displayed alongside the outcome data from the effectiveness evaluation. This could be either in main either in the main manuscript, journal companion paper or supplementary material.

Item 13. How can the resource use evidence be used to support public health practice and research?

Firstly, the disaggregated resource use evidence can be used to support immediate local government's public health decisions about the likely resource implications of implementing an PA/SB intervention. This is appropriate in the UK, since trial-derived resource use evidence is scarce,

yet there is the demand for this data amongst public health decision makers who operate of short-term budgets. Secondly, the resource use data can feed into a full trial- and/or model- based economic evaluation. This can be done by the health economist carrying out the second step in costing after resource use measurement, which involves the assignment of unit costs to resource use items. Lastly, researchers involved in the development and design of interventions can use the resource use data to explore trends in levels of resource use and outcomes. The APEASE framework is a key tool from the behaviour change and implementation science literature (Michie et al. 2014). APEASE recognises that intervention design is more than effectiveness and recognises that the social context in which an intervention operates is also important. More specifically, the APEASE framework is designed to encourage researches to collect evidence on other factors such as affordability, practicability, acceptability and equity. Resource use data collected in accordance to this SOP can provide evidence to inform these specific factors.

Chapter 7: Conclusions

7.1. Wider implications

7.1.1. Implications for policy

The framework supports the generation of resource use (economic) data. This type of data is in high demand amongst public sector decision makers within many countries (Frew et al., 2018). There is widespread interest in economic data with the number of trials collecting economic data increasing (Ramsey et al., 2015). In the public sector, year on year the demand for cost data continues to grow (Curtis and Burns, 2018). Reasons for this include the fact that in a number of countries public sector budgets are tighter and healthcare demand is increasing (Weatherly et al., 2014). In the UK, it is reported that Public Health Directors are 'hungry' for economic evidence on the short-term economic impact of the preventative interventions they commission (Willmott, Womack, Hollingworth, & Campbell, 2016). In the UK, there has been increased interest in improving costing methods, with the National Cost Collection (NCC) programme moving away from reference costs to patient-level costing (NHS improvement, 2019). Trueman and Anokye (2013) refer to CUA as a 'powerful common currency tool' for decision-makers, but point out that the level of detail provided in CCAs is more likely to be desirable to local level public health commissioners and non-health decision-makers. Overall, real-world decisions around behaviour change interventions are not just about effectiveness and cost-effectiveness, but about if the intervention is affordable, equitable, acceptable, practical and safe (Michie et al., 2014). Good economic evaluations depend on the collection and presentation of robust data. Outside the academic world, public and third sector agencies who have to make inevitable resource allocation decisions are drawing on other methodologies (e.g. SROI and break-even analysis) in order to understand value. The SROI methodology is growing rapidly, despite the approach's academic validity being questioned (Fujiwara, 2015). Similarly, there is concern about the fact these novel economic methodologies have not undergone the same peer review process as happens with academic papers (Svistak and Pritchard, 2018). Overall, the framework enables the collection and presentation of resource use and quality of life evidence to become more commonplace. Policymakers can use the evidence generated through the framework, to inform their unavoidable resource allocation decisions, which they make on behalf of the population. In turn, the evidence has the potential to make the policymakers more accountable for their decisions, which encourages transparency.

7.1.2. Recommendations for future research

7.1.2.1. Piloting the framework with other researchers

One way to refine the frameowkr further could be to conduct an RCT of the framework itself. One randomised group of trialists with no or limited experience of health economics, could be asked to apply the refined framework to a future upcoming trial, while a control group could be asked to conduct their trials as usual. A questionnaire could be disseminated to quantify how many and which items on the framework they adopted for the group using the framework. In addition, both the trialists using the framework and the control group could be asked about whether they used any other tools to support them in collecting and presenting resource use data. Focus groups and

interviews could also be done to see how practical the trialists found the framework and the presentation of the resource use data in the publications could be compared. The content of the publications could also be reviewed to see if the trialists applying the framework have reported resource use quantitatively alongside their effectiveness data. Further piloting can be done if only minor adjustments are required, or if more major adjustments are needed, a Delphi study could be done with a larger sample of trialists from around the UK, to help reconsider the design of the framework.

7.1.2.2. Measurement tools for resource use

More research is needed to develop the tools used to measure resource use since there are many limitations reported for the existing tools. Length of the questionnaire can influence uptake. This was suggested as a reason for the lack of update of the annotated patient cost questionnaire (Thorn et al., 2018, Thompson and Wordsworth, 2001). GP records have been found to provide more reliable estimates on visits to health professionals in GP surgeries (e.g. the GP and Practice Nurse) compared to patient-recall (Byford et al., 2007). However, in the same study it was found that GP records generated less reliable estimates on visits to primary and community care health professionals outside of GP surgeries, and secondary care services. Although there are concerns around patient recall bias, one review found that there was good agreement between patient self-report data and medical records (Ridyard and Hughes, 2015). Self-report methods are still the preferred option for resource use data in economic evaluations in England (especially if secondary care is not a major cost) due to the time-consuming data extraction periods and the data sharing agreements associated with electronic datasets (Franklin and Thorn, 2019). An additional, problem with patient self-report measures is that they have been found to be prone to recall bias (Jessep et al., 2009) and patients do not reportedly know which specific type of health professionals they have seen (Thorn et al., 2018). One study found that patients did not know what the term ‘Practice Nurse’ means, the authors argue that definitions need to be provided alongside this (Byford et al., 2007). A strength of the healthcare utilisation questionnaires used in the SLaMM and Co-PARS trial was that they were modified to include definitions for outpatient, inpatient and day case hospital visits. Future measurement tools should also include definitions to help participants understand the type of primary or secondary healthcare you are interested in.

7.1.2.3. Microcosting

Although microcosting is the preferred approach to costing, analysts have typically combined different levels of costing and thus different levels of accuracy within a single study. This is due to the lack of resource use data available (Mogyorosy and Smith, 2005). The authors argue that bottom-up approaches are likely to be more accurate for complex interventions since the input mix of resources is complex. Future studies should continue to collect more accurate resource data on the interventions they are delivering. Collecting just resource quantities may reduce the expensive, high researcher burden and practical challenges of microcosting (Mogyorosy and Smith, 2005). It has been reported in a key review that interventions are poorly described in the literature (Hoffmann et al., 2014). The review found the materials used in the delivery of an intervention are the most commonly omitted piece of information when researchers describe their interventions. This makes it challenging for others to replicate this intervention in their own setting. That is to say,

until valid and reliable standardised microcosting tools are available, data collection tools may need to be tailored to the intervention staff's preferences in order achieve good completion rates (Hughes et al., 2016). In addition, a recommendation from this thesis is that completion of the objective absenteeism data should be assessed prior to the start of the trial, so additional self-report measures can be included whereby participants are asked to report the number of sick days they have had over the desired time point.

7.1.2.4. Quality of life measurement tools

The EQ-5D tool was the most commonly used tool to measure QALYs in the studies reviewed in Chapter 2, and in the wider literature. Nonetheless, the EQ-5D only captures the functional health of an individual. Future studies could use other recently developed quality of life tools such as the ICECAP-A (Al-Janabi et al., 2012, Al-Janabi et al., 2013, Flynn et al., 2015), which has been designed to capture capability in a broader sense, beyond functional health. Another solution is for analysts to agree on a tool which crosswalks between PA outcomes and a summary tool like the EQ-5D. There is currently a mapping database of studies that map the EQ-5D tool to other outcomes measures (Dakin et al., 2018). No studies on the database have mapped a PA specific tool to the EQ-5D; future research should address this gap.

7.2. Interdisciplinary researcher

This PhD project provided me with the opportunity to train in a new discipline, health economics. It also provided me with the opportunity to build on my existing discipline-specific knowledge in public health and anthropology. In particular, I believe this PhD project has enabled me to develop the practical skills and knowledge on how to identify and measure resource use relevant to PA and SB trials. More specifically, I have improved my skills in working with different types of stakeholders (e.g. staff from the local authority, leisure centre, workplace and the participants) as well as researchers with different disciplinary backgrounds (public health researchers, exercise science experts, physiologists, psychologists, trialists and health economists). The meetings and informal conversations I had with these various stakeholders and researchers made me aware of how important these relationships are for understanding and mapping out what resource use is relevant and for gaining an insight on how resource use can be measured and quantified in PA and SB trials. Furthermore, over the last three years, I feel I have becoming increasingly more confident and skilled in articulating health economic, public health and anthropology concepts. This skill developed as a result of the range of multi-stakeholder and multidisciplinary meetings I arranged throughout this PhD. This is also evident through the definitions I provide in my SOP which are intended to support multidisciplinary learning. Throughout this PhD, I handled both self-report resource use data (e.g. intervention costs, healthcare utilisation, participant out-of-pocket and productivity loss) and electronic absenteeism data (e.g. company records). I believe this PhD has strengthened my awareness and knowledge on the strengths and limitations of using different data collection approaches in the measurement of resource use for PA and SB trials. My training in health economics also raised my awareness of the concept opportunity cost. This is a concept that is not commonly discussed in the public health literature which is surprising since it is relevant to the types of decisions and trade-offs national and local public health decisions-makers frequently made on behalf of the others.

7.3. Novel contributions of the thesis

There are four novel contributions of this thesis, these include:

- My development as a true interdisciplinary researcher. More specifically, I have trained in health economics so as I can make an important interdisciplinary contribution to the field of public health economics. I have demonstrated how my knowledge in public health and anthropology has contributed to the studying of complexity within PA and SB trials.
- My systematic review provides an update on how analysts are addressing the four methodological challenges associated with economic evaluations of individual-level PA/ SB interventions.
- I present a SOP to support early career health economists to identify, measure and present resource use in a systematic way alongside individual-level PA and SB trials. Presently, the trial unit I work at does not have a SOP for identifying and measuring resource use data for complex public health trials, including individual-level PA and SB trials. I intend to apply my SOP to future PA and SB trials and disseminate it amongst other early career health economists.
- The SOP can be used as an interface to improve understanding and communication for multidisciplinary trial teams. In particular, the SOP can be an interface between the health economist and members of the trail team who do not have expertise in health economics.

7.4. Conclusion

Trials evaluating the impact of PA and SB interventions rarely collect and present economic data alongside their effectiveness results. The overarching aim of this PhD was to develop a multidisciplinary and pragmatic framework to support researchers carrying out trial-based economic evaluations for individual-level PA and SB interventions. Three studies, a systematic review and two PA and SB economic evaluations, were carried out to specifically address this aim. A unique feature of these studies was the reflective content embedded throughout. The reflective content describes in detail the interdisciplinary actions I took in order to develop and implement a multidisciplinary and pragmatic framework. Methods of narrative synthesis were used to organise the reflective content from this thesis in order to consider how the practicability of the framework could be improved. The narrative synthesis shed light on how the framework could be refined to support early career health economists to tackle on the ground some of the complexity involved in the identification and measurement of resource use in PA and SB trials. In order to promote multidisciplinary working in PA and SB trials, the final framework is presented in the form of a SOP which can be read by all key members of the trial team. In particular, the SOP can act as an interface between the health economist and other trial team members in order to improve communication across the multiple disciplines in which the fields of PA and SB cut across. Importantly, SOP is expected to improve the practicability, consistency, transparency and efficiency of the identification and measurement of resource use alongside individual-level PA and SB trials.

References

References

ABU-OMAR, K., RÜTTEN, A., BURLACU, I., SCHÄTZLEIN, V., MESSING, S. & SUHRCKE, M. 2017. The cost-effectiveness of physical activity interventions: A systematic review of reviews. *Preventive Medicine Reports,* 8**,** 72-78.

AINSWORTH, B. E., HASKELL, W. L., HERRMANN, S. D., MECKES, N., BASSETT, D. R., JR., TUDOR-LOCKE, C., GREER, J. L., VEZINA, J., WHITT-GLOVER, M. C. & LEON, A. S. 2011. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc,* 43**,** 1575-81.

AL-JANABI, H., FLYNN, T. N. & COAST, J. 2012. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research,* 21**,** 167-176.

AL-JANABI, H., PETERS, T. J., BRAZIER, J., BRYAN, S., FLYNN, T. N., CLEMENS, S., MOODY, A. & COAST, J. 2013. An investigation of the construct validity of the ICECAP-A capability measure. *Quality of Life Research,* 22**,** 1831-1840.

ALAYLI-GOEBBELS, A. F. G., EVERS, S. M. A. A., ALEXEEVA, D., AMENT, A. J. H. A., DE VRIES, N. K., TILLY, J. C. & SEVERENS, J. L. 2014. A review of economic evaluations of behavior change interventions: setting an agenda for research methods and practice. *Journal of Public Health,* 36**,** 336-344.

ALLEN, L. 2015. *Why cutting spending on public health is a false economy. The Conversation.* [Online]. The Conversation,: The Conversation,. Available: https://theconversation.com/why-cutting-spending-on-public-health-is-a-false-economy-51159 [Accessed 22nd August 2019].

ANDERSEN, L. B., MOTA, J. & DI PIETRO, L. 2016. Update on the global pandemic of physical inactivity. *The Lancet,* 388**,** 1255-1256.

ANDERSON, R. 2010. Systematic reviews of economic evaluations: utility or futility? *Health Economics,* 19**,** 350-364.

ANOKYE, N. K., FOX-RUSHBY, J., SANGHERA, S., COOK, D. G., LIMB, E., FURNESS, C., KERRY, S. M., VICTOR, C. R., ILIFFE, S., USSHER, M., WHINCUP, P. H., EKELUND, U., DEWILDE, S. & HARRIS, T. 2018. Short-term and long-term cost-effectiveness of a pedometer-based exercise intervention in primary care: a within-trial analysis and beyond-trial modelling. *BMJ Open,* 8**,** e021978.

ANOKYE, N. K., LORD, J. & FOX-RUSHBY, J. 2012. National Institute for Health and Clinical Excellence: Public Health Intervention Guidance on Physical Activity – Brief advice for adults in primary care: Economic Analysis. London: National Institute for Health and Clinical Excellence,.

ANOKYE, N. K., LORD, J. & FOX-RUSHBY, J. 2014. Is brief advice in primary care a cost-effective way to promote physical activity? *British Journal of Sports Medicine,* 48**,** 202-206.

ANOKYE, N. K., TRUEMAN, P., GREEN, C., PAVEY, T., HILLSDON, M. & TAYLOR, R. 2011. The cost-effectiveness of exercise referral schemes. *BMC Public Health,* 11**,** 954.

BARISIC, A., LEATHERDALE, S. T. & KREIGER, N. 2011. Importance of frequency, intensity, time and type (FITT) in physical activity assessment for epidemiological research. *Canadian journal of public health = Revue canadienne de sante publique,* 102**,** 174-175.

BEECHAM, J. & KNAPP, M. 2001. *12 Costing psychiatric interventions* [Online]. Available: https://www.researchgate.net/publication/242359523_12_Costing_psychiatric_interventions [Accessed 17th May 2019].

BLAIR, S. N. 2009. Physical inactivity: the biggest public health problem of the 21st century. *British Journal of Sports Medicine,* 43**,** 1.

BLATCH-JONES, A. J., PEK, W., KIRKPATRICK, E. & ASHTON-KEY, M. 2018. Role of feasibility and pilot studies in randomised controlled trials: a cross-sectional study. *BMJ Open,* 8**,** e022233.

BOADWAY, R. W. & BRUCE, N. 1984. *Welfare Economics,* Oxford, England, Basil Blackwell.

BOBERSKA, M., SZCZUKA, Z., KRUK, M., KNOLL, N., KELLER, J., HOHL, D. H. & LUSZCZYNSKA, A. 2018. Sedentary behaviours and health-related quality of life. A systematic review and meta-analysis. *Health Psychol Rev,* 12**,** 195-210.

BOEHLER, C. E. H., MILTON, K. E., BULL, F. C. & FOX-RUSHBY, J. A. 2011. The cost of changing physical activity behaviour: evidence from a "physical activity pathway" in the primary care setting. *BMC Public Health,* 11**,** 370.

BOMMER, C., HEESEMANN, E., SAGALOVA, V., MANNE-GOEHLER, J., ATUN, R., BÄRNIGHAUSEN, T. & VOLLMER, S. 2017. The global economic burden of diabetes in

adults aged 20–79 years: a cost-of-illness study. *The Lancet Diabetes & Endocrinology,* 5**,** 423-430.
BOOTH, F. W., LAYE, M. J., LEES, S. J., RECTOR, R. S. & THYFAULT, J. P. 2008. Reduced physical activity and risk of chronic disease: the biology behind the consequences. *European Journal of Applied Physiology,* 102**,** 381-390.
BRIGGS, A. 2000. Economic evaluation and clinical trials: size matters. *BMJ (Clinical research ed.),* 321**,** 1362-1363.
BRIGGS, A. S., SCULPHER, M. J. & CLAXTON, K. 2006. *Decision Modelling for Health Economic Evaluation,* Oxford, Oxford University Press.
BROUWER, W. B. F. & KOOPMANSCHAP, M. A. 2000. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *Journal of Health Economics,* 19**,** 439-459.
BRYAN, S. & WILLIAMS, I. 2014. Adoption of New Technologies, Using Economic Evaluation. *In:* CULYER, A. J. (ed.) *Encyclopedia of Health Economics.* Oxford: Elsevier.
BUCKLEY, B. J. R., THIJSSEN, D. H. J., MURPHY, R. C., GRAVES, L. E. F., WHYTE, G., GILLISON, F., CRONE, D., WILSON, P. M., HINDLEY, D. & WATSON, P. M. 2019. Preliminary effects and acceptability of a co-produced physical activity referral intervention. *Health Education Journal***,** 0017896919853322.
BUCKLEY, B. J. R., THIJSSEN, D. H. J., MURPHY, R. C., GRAVES, L. E. F., WHYTE, G., GILLISON, F., CRONE, D., WILSON, P. M. & WATSON, P. M. 2018. Making a move in exercise referral: co-development of a physical activity referral scheme. *Journal of Public Health,* 40**,** e586-e593.
BUCKLEY, J. P., HEDGE, A., YATES, T., COPELAND, R. J., LOOSEMORE, M., HAMER, M., BRADLEY, G. & DUNSTAN, D. W. 2015. The sedentary office: a growing case for change towards better health and productivity. Expert statement commissioned by Public Health England and the Active Working Community Interest Company. *British Journal of Sports Medicine*.
BUSSE, R., SCHREYÖGG, J. & SMITH, P. C. 2008. Variability in healthcare treatment costs amongst nine EU countries – results from the HealthBASKET project. *Health Economics,* 17**,** S1-S8.
BYFORD, S., LEESE, M., KNAPP, M., SEIVEWRIGHT, H., CAMERON, S., JONES, V., DAVIDSON, K. & TYRER, P. 2007. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health Economics,* 16**,** 531-536.
CAMPBELL, F., HOLMES, M., EVERSON-HOCK, E., DAVIS, S., BUCKLEY WOODS, H. & ANOKYE, N. 2015a. A systematic review and economic evaluation of exercise referral schemes in primary care: a short report. *Health Technology Assessment (Winchester, England),* 19**,** 1-110.
CAMPBELL, F., HOLMES, M., EVERSON-HOCK, E., DAVIS, S., BUCKLEY WOODS, H. & ANOKYE, N. 2015b. A systematic review and economic evaluation of exercise referral schemes in primary care: a short report. 19.
CARSON, V., HUNTER, S., KUZIK, N., GRAY, C. E., POITRAS, V. J., CHAPUT, J. P., SAUNDERS, T. J., KATZMARZYK, P. T., OKELY, A. D., CONNOR GORBER, S., KHO, M. E., SAMPSON, M., LEE, H. & TREMBLAY, M. S. 2016. Systematic review of sedentary behaviour and health indicators in school-aged children and youth: an update. *Appl Physiol Nutr Metab,* 41**,** S240-65.
CASONATTO, J., GOESSLER, K. F., CORNELISSEN, V. A., CARDOSO, J. R. & POLITO, M. D. 2016. The blood pressure-lowering effect of a single bout of resistance exercise: A systematic review and meta-analysis of randomised controlled trials. *Eur J Prev Cardiol,* 23**,** 1700-1714.
CASPERSEN, C. J., POWELL, K. E. & CHRISTENSON, G. M. 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports (Washington, D.C. : 1974),* 100**,** 126-131.
CHARLES, J. M., EDWARDS, R. T., BYWATER, T. & HUTCHINGS, J. 2013. Micro-Costing in Public Health Economics: Steps Towards a Standardized Framework, Using the Incredible Years Toddler Parenting Program as a Worked Example. *Prevention Science,* 14**,** 377-389.
CHEKROUD, S. R., GUEORGUIEVA, R., ZHEUTLIN, A. B., PAULUS, M., KRUMHOLZ, H. M., KRYSTAL, J. H. & CHEKROUD, A. M. 2018. Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study. *The lancet. Psychiatry,* 5**,** 739-746.
CHU, A. H. Y., NG, S. H. X., TAN, C. S., WIN, A. M., KOH, D. & MÜLLER-RIEMENSCHNEIDER, F. 2016. A systematic review and meta-analysis of workplace intervention strategies to reduce sedentary time in white-collar workers. *Obesity Reviews,* 17**,** 467-481.

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**,** 1-503, v-vi.

CLEMES, S. A., O'CONNELL, S. E. & EDWARDSON, C. L. 2014. Office workers' objectively measured sedentary behavior and physical activity during and outside working hours. *J Occup Environ Med,* 56**,** 298-303.

COMET. 2020. *Core Outcome Measures in Effectiveness Trials* [Online]. Available: http://www.comet-initiative.org/ [Accessed 05.07.20].

COOKSON, R., MIRELMAN, A. J., GRIFFIN, S., ASARIA, M., DAWKINS, B., NORHEIM, O. F., VERGUET, S. & J. CULYER, A. 2017. Using Cost-Effectiveness Analysis to Address Health Equity Concerns. *Value in Health,* 20**,** 206-212.

CRAIG, P., DIEPPE, P., MACINTYRE, S., MICHIE, S., NAZARETH, I., PETTICREW, M. & MEDICAL RESEARCH COUNCIL, G. 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ (Clinical research ed.),* 337**,** a1655-a1655.

CULYER, A. J., COOKSON, R. & CLAXTON, K. 2012. *The Humble Economist,* York, York Publishing Services Ltd.

CURTIS, L. A. & BURNS, A. 2018. *Unit Costs of Health and Social Care 2018,* Canterbury, PSSRU, University of Kent.

CYLUS, J., PAPANICOLAS, I. & SMITH, P. C. 2016. A framework for thinking about health system efficiency. *Health System Efficiency***,** 1.

DAHLGREN, G. & WHITEHEAD, M. 1991. Policies and strategies to promote social equity in health (Mimeo). Stockholm, Sweden: Insititute for Future Studies,.

DAKIN, H., ABEL, L., BURNS, R. & YANG, Y. 2018. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health and Quality of Life Outcomes,* 16**,** 31.

DAVIS, J. C., VERHAGEN, E., BRYAN, S., LIU-AMBROSE, T., BORLAND, J., BUCHNER, D., HENDRIKS, M. R., WEILER, R., MORROW, J. R., VAN MECHELEN, W., BLAIR, S. N., PRATT, M., WINDT, J., AL-TUNAIJI, H., MACRI, E. & KHAN, K. M. 2014. 2014 Consensus Statement from the first Economics of Physical Inactivity Consensus (EPIC) Conference (Vancouver). *British Journal of Sports Medicine,* 48**,** 947-951.

DE VRIES, N. M., STAAL, J. B., TEERENSTRA, S., ADANG, E. M. M., RIKKERT, M. G. M. O. & NIJHUIS-VAN DER SANDEN, M. W. G. 2013. Physiotherapy to improve physical activity in community-dwelling older adults with mobility problems (Coach2Move): study protocol for a randomized controlled trial. *Trials,* 14**,** 434.

DE VRIES, N. M., STAAL, J. B., VAN DER WEES, P. J., ADANG, E. M., AKKERMANS, R., OLDE RIKKERT, M. G. & NIJHUIS-VAN DER SANDEN, M. W. 2016. Patient-centred physical therapy is (cost-) effective in increasing physical activity and reducing frailty in older adults with mobility problems: a randomized controlled trial with 6 months follow-up. *J Cachexia Sarcopenia Muscle,* 7**,** 422-35.

DEPARTMENT FOR COMMUNITIES AND LOCAL GOVERNMENT 2015. English indices of deprivation. London: Department for Communities and Local Government,.

DEPARTMENT OF HEALTH 2004. At least five a week: Evidence on the impact of physical activity and its relationship to health. A report from the Chief Medical Officer. London: HMSO.

DEPARTMENT OF HEALTH AND HUMAN SERVICES. 2008. *Physical Activity Guidelines for Americans.* [Online]. Available: https://health.gov/paguidelines/2008/ [Accessed 20th August 2019].

DEPARTMENT OF HEALTH AND SOCIAL CARE 2011. *Start Active, Stay Active: A Report on Physical Activity for Health from the Four Home Countries' Chief Medical Officers,* London, Department of Health.

DING, D., LAWSON, K. D., KOLBE-ALEXANDER, T. L., FINKELSTEIN, E. A., KATZMARZYK, P. T., VAN MECHELEN, W. & PRATT, M. 2016. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet,* 388**,** 1311-1324.

DOLAN, P. 1997. Modeling valuations for EuroQol health states. 35**,** 1095-1108.

DONALDSON, C., BIRCH, S. & GAFNI, A. 2002. The distribution problem in economic evaluation: income and the valuation of costs and consequences of health care programmes. *Health Economics,* 11**,** 55-70.

DRITSAKI, M., GRAY, A., PETROU, S., DUTTON, S., LAMB, S. E. & THORN, J. C. 2018. Current UK Practices on Health Economics Analysis Plans (HEAPs): Are We Using Heaps of Them? *Pharmacoeconomics,* 36**,** 253-257.

DRUMMOND, M., KLAXTON, K., STODDART, G. & TORRANCE, G. 2015a. *Methods for the Economic Evaluation of Health Care Programmes,* Oxford, Oxford University Press.

DRUMMOND, M., SCHULPHER, M. J., CLAXTON, K., STODDART, G. L. & TORRANCE, G. W. 2015b. *Methods for the Economic Evaluation of Health Care Programmes,* Oxford, Oxford University Press.
DUGDILL, L., GRAHAM, R. C. & MCNAIR, F. 2005. Exercise referral: the public health panacea for physical activity promotion? A critical perspective of exercise referral schemes; their development and evaluation. *Ergonomics,* 48**,** 1390-1410.
DUNSTAN, D. W., SALMON, J., OWEN, N., ARMSTRONG, T., ZIMMET, P. Z., WELBORN, T. A., CAMERON, A. J., DWYER, T., JOLLEY, D. & SHAW, J. E. 2004. Physical activity and television viewing in relation to risk of undiagnosed abnormal glucose metabolism in adults. *Diabetes Care,* 27**,** 2603-9.
EDWARDS, R. T., CHARLES, J. M. & LLOYD-WILLIAMS, H. 2013a. 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,* 13.
EDWARDS, R. T., LINCK, P., HOUNSOME, N., RAISANEN, L., WILLIAMS, N., MOORE, L. & MURPHY, S. 2013b. Cost-effectiveness of a national exercise referral programme for primary care patients in Wales: results of a randomised controlled trial. *Bmc Public Health,* 13.
EDWARDS, R. T. & MCINTOSH, E. 2019. *Applied Health Economics for Public Health Practice and Research,* Oxford, UK, Oxford University Press.
EDWARDSON, C. L., YATES, T., BIDDLE, S. J. H., DAVIES, M. J., DUNSTAN, D. W., ESLIGER, D. W., GRAY, L. J., JACKSON, B., O'CONNELL, S. E., WAHEED, G. & MUNIR, F. 2018. Effectiveness of the Stand More AT (SMArT) Work intervention: cluster randomised controlled trial. *BMJ,* 363**,** k3870.
EEKHOUT, I., DE VET, H. C., TWISK, J. W., BRAND, J. P., DE BOER, M. R. & HEYMANS, M. W. 2014. Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology,* 67**,** 335-42.
EKELUND, U., STEENE-JOHANNESSEN, J., BROWN, W. J., FAGERLAND, M. W., OWEN, N., POWELL, K. E., BAUMAN, A. & LEE, I. M. 2016. Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. *Lancet,* 388**,** 1302-10.
ELLEY, C. R., GARRETT, S., ROSE, S. B., O'DEA, D., LAWTON, B. A., MOYES, S. A. & DOWELL, A. C. 2011. Cost-effectiveness of exercise on prescription with telephone support among women in general practice over 2 years. *British Journal of Sports Medicine,* 45**,** 1223-9.
FLYNN, T. N., HUYNH, E., PETERS, T. J., AL-JANABI, H., CLEMENS, S., MOODY, A. & COAST, J. 2015. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Economics,* 24**,** 258-269.
FRANKLIN, M., LOMAS, J., WALKER, S. & YOUNG, T. 2019. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. *Pharmacoeconomics*.
FRANKLIN, M. & THORN, J. 2019. Self-reported and routinely collected electronic healthcare resource-use data for trial-based economic evaluations: the current state of play in England and considerations for the future. *BMC Medical Research Methodology,* 19**,** 8.
FREW, E. & BREHENY, K. 2019. Methods for public health economic evaluation: A Delphi survey of decision makers in English and Welsh local government. *Health Economics,* 28**,** 1052-1063.
FREW, E., PLATT, M., BROWN, H., GOODMAN, A., SEIDLER, E., SPETZ, J. & MCINTYRE, D. 2018. Building an international health economics teaching network. *Health Economics,* 27**,** 919-922.
FRICK, K. D. 2009. Microcosting quantity data collection methods. *Medical Care,* 47**,** S76-81.
FUJIWARA, D. 2015. The Seven Principle Problems of SROI. London, UK: Simetrica.
FÜZÉKI, E., ENGEROFF, T. & BANZER, W. 2017. Health Benefits of Light-Intensity Physical Activity: A Systematic Review of Accelerometer Data of the National Health and Nutrition Examination Survey (NHANES). *Sports Medicine,* 47**,** 1769-1793.
GAO, L., FLEGO, A., DUNSTAN, D. W., WINKLER, E. A. H., HEALY, G. N., EAKIN, E. G., WILLENBERG, L., OWEN, N., LAMONTAGNE, A. D., LAL, A., WIESNER, G. H., HADGRAFT, N. T. & MOODIE, M. L. 2018. Economic evaluation of a randomized controlled trial of an intervention to reduce office workers' sitting time: the "Stand Up Victoria" trial. *Scandinavian Journal of Work, Environment & Health***,** 503-511.
GARDNER, B., SMITH, L., LORENCATTO, F., HAMER, M. & BIDDLE, S. J. H. 2016. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. *Health Psychology Review,* 10**,** 89-112.

GC, V., WILSON, E. C. F., SUHRCKE, M., HARDEMAN, W. & SUTTON, S. 2016. Are brief interventions to increase physical activity cost-effective? A systematic review. *British Journal Of Sports Medicine,* 50**,** 408-417.
GIBSON-MOORE, H. 2019. UK Chief Medical Officers' physical activity guidelines 2019: What's new and how can we get people more active? *Nutrition Bulletin,* 44**,** 320-328.
GLICK, H. A., DOSHI, J. A., SONNAD, S. S. & POLSKY, D. 2014. *Economic Evaluation in Clinical Trials*, Oxford University Press.
GRANT, S., MAYO-WILSON, E., MONTGOMERY, P., MACDONALD, G., MICHIE, S., HOPEWELL, S. & MOHER, D. 2018. CONSORT-SPI 2018 Explanation and Elaboration: guidance for reporting social and psychological intervention trials. *Trials,* 19**,** 406.
GRAY, A. M., CLARKE, P. M., WOLSTENHOLME, J. L. & WORDSWORTH, S. 2012. *Applied Methods of Cost-effectiveness Analysis in Health Care,* Oxford, Oxford University Press.
GREENBERG, D., ROSEN, A. B., OLCHANSKI, N. V., STONE, P. W., NADAI, J. & NEUMANN, P. J. 2004. Delays in publication of cost utility analyses conducted alongside clinical trials: registry analysis. *BMJ (Clinical Research ed.),* 328**,** 1536-7.
HAINES, A., KURUVILLA, S. & BORCHERT, M. 2004. Bridging the implementation gap between knowledge and action for health. *Bull World Health Organ,* 82**,** 724-31; discussion 732.
HALLAL, P. C., ANDERSEN, L. B., BULL, F. C., GUTHOLD, R., HASKELL, W. & EKELUND, U. 2012. Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet,* 380**,** 247-257.
HAMER, M., COOMBS, N. & STAMATAKIS, E. 2014. Associations between objectively assessed and self-reported sedentary time with mental health in adults: an analysis of data from the Health Survey for England. *BMJ Open,* 4**,** e004580.
HAMILTON, M. T., HEALY, G. N., DUNSTAN, D. W., ZDERIC, T. W. & OWEN, N. 2008. Too Little Exercise and Too Much Sitting: Inactivity Physiology and the Need for New Recommendations on Sedentary Behavior. *Current cardiovascular risk reports,* 2**,** 292-298.
HARRIS, T., KERRY, S. M., LIMB, E. S., FURNESS, C., WAHLICH, C. & VICTOR, C. R. 2018. Physical activity levels in adults and older adults 3-4 years after pedometer-based walking interventions: long-term follow-up of participants from two randomised controlled trials in UK primary care. 15**,** e1002526.
HEALY, G. N., EAKIN, E. G., LAMONTAGNE, A. D., OWEN, N., WINKLER, E. A., WIESNER, G., GUNNING, L., NEUHAUS, M., LAWLER, S., FJELDSOE, B. S. & DUNSTAN, D. W. 2013. Reducing sitting time in office workers: short-term efficacy of a multicomponent intervention. *Prev Med,* 57**,** 43-8.
HEATH, G. W., PARRA, D. C., SARMIENTO, O. L., ANDERSEN, L. B., OWEN, N., GOENKA, S., MONTES, F. & BROWNSON, R. C. 2012. Evidence-based intervention in physical activity: lessons from around the world. *The Lancet,* 380**,** 272-281.
HILL, S. R., VALE, L., HUNTER, D., HENDERSON, E. & OLUBOYEDE, Y. 2017. Economic evaluations of alcohol prevention interventions: Is the evidence sufficient? A review of methodological challenges. *Health Policy,* 121**,** 1249-1262.
HM TREASURY. 2019. *GDP deflators at market prices, and money GDP: March 2019 (Spring Statement)* [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/786181/GDP_Deflators_Spring_Statement_2019_update.csv/preview [Accessed 18.06.2019].
HOFFMANN, T. C., GLASZIOU, P. P., BOUTRON, I., MILNE, R., PERERA, R., MOHER, D., ALTMAN, D. G., BARBOUR, V., MACDONALD, H., JOHNSTON, M., LAMB, S. E., DIXON-WOODS, M., MCCULLOCH, P., WYATT, J. C., CHAN, A.-W. & MICHIE, S. 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ : British Medical Journal,* 348.
HU, G., JOUSILAHTI, P., BORODULIN, K., BARENGO, N. C., LAKKA, T. A., NISSINEN, A. & TUOMILEHTO, J. 2007. Occupational, commuting and leisure-time physical activity in relation to coronary heart disease among middle-aged Finnish men and women. *Atherosclerosis,* 194**,** 490-497.
HU, G., QIAO, Q., SILVENTOINEN, K., ERIKSSON, J. G., JOUSILAHTI, P., LINDSTROM, J., VALLE, T. T., NISSINEN, A. & TUOMILEHTO, J. 2003. Occupational, commuting, and leisure-time physical activity in relation to risk for Type 2 diabetes in middle-aged Finnish men and women. *Diabetologia,* 46**,** 322-9.
HU, G., SARTI, C., JOUSILAHTI, P., SILVENTOINEN, K., BARENGO, N. C. & TUOMILEHTO, J. 2005. Leisure time, occupational, and commuting physical activity and the risk of stroke. *Stroke,* 36**,** 1994-1999.
HUGHES, D., CHARLES, J., DAWOUD, D., EDWARDS, R. T., HOLMES, E., JONES, C., PARHAM, P., PLUMPTON, C., RIDYARD, C., LLOYD-WILLIAMS, H., WOOD, E. & YEO,

S. T. 2016. Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches. *Pharmacoeconomics,* 34**,** 447-461.
HUPIN, D., ROCHE, F., GREMEAUX, V., CHATARD, J. C., ORIOL, M., GASPOZ, J. M., BARTHÉLÉMY, J. C. & EDOUARD, P. 2015. Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: a systematic review and meta-analysis. *Br J Sports Med,* 49**,** 1262-7.
HUSEREAU, D., DRUMMOND, M., PETROU, S., CARSWELL, C., MOHER, D., GREENBERG, D., AUGUSTOVSKI, F., BRIGGS, A. H., MAUSKOPF, J. & LODER, E. 2013. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ : British Medical Journal,* 346.
ILIFFE, S., KENDRICK, D., MORRIS, R., MASUD, T., GAGE, H. & SKELTON, D. 2014a. Multicentre cluster randomised trial comparing a community group exercise programme and home-based exercise with usual care for people aged 65 years and over in primary care. 18.
ILIFFE, S., KENDRICK, D., MORRIS, R., MASUD, T., GAGE, H., SKELTON, D., DINAN, S., BOWLING, A., GRIFFIN, M., HAWORTH, D., SWANWICK, G., CARPENTER, H., KUMAR, A., STEVENS, Z., GAWLER, S., BARLOW, C., COOK, J. & BELCHER, C. 2014b. Multicentre cluster randomised trial comparing a community group exercise programme and home-based exercise with usual care for people aged 65 years and over in primary care. *Health Technol Assess,* 18**,** vii-xxvii, 1-105.
INTERNATIONAL SOCIETY FOR PHYSICAL ACTIVITY AND HEALTH 2012. Investments that Work for Physical Activity. *British Journal of Sports Medicine,* 46**,** 709.
ISAACS, A. J., CRITCHLEY, J. A., TAI, S. S., BUCKINGHAM, K., WESTLEY, D., HARRIDGE, S. D., SMITH, C. & GOTTLIEB, J. M. 2007. Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only. *Health Technol Assess,* 11**,** 1-165, iii-iv.
ISPOR. 2017. *Minimal Clinically Important Difference in EQ-5D: We Can Calculate it – But Does That Mean We Should?* [Online]. Available: https://www.ispor.org/docs/default-source/presentations/1066.pdf?sfvrsn=25feffd6_1 [Accessed 04.07.2019].
ISPOR. 2018. *Top 10 HEOR Trends* [Online]. Available: https://www.ispor.org/docs/default-source/publications/top10trends.pdf?sfvrsn=e6052ae7_2 [Accessed 30.08.2018].
ISPOR. 2019. *Pharmacoeconomic Guidelines Around The World* [Online]. Available: https://tools.ispor.org/peguidelines/ [Accessed 24.04.2019].
JESSEP, S. A., WALSH, N. E., RATCLIFFE, J. & HURLEY, M. V. 2009. Long-term clinical benefits and costs of an integrated rehabilitation programme compared with outpatient physiotherapy for chronic knee pain. *Physiotherapy,* 95**,** 94-102.
KATZMARZYK, P. T., CHURCH, T. S., CRAIG, C. L. & BOUCHARD, C. 2009. Sitting time and mortality from all causes, cardiovascular disease, and cancer. *Med Sci Sports Exerc,* 41**,** 998-1005.
KELLY, M. P., MCDAID, D., LUDBROOK, A. & POWELL, J. E. 2005. *Economic appraisal of public health interventions.* [Online]. Health Development Agency,. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.3820&rep=rep1&type=pdf [Accessed 10.05.2018].
KIGOZI, J., JOWETT, S., LEWIS, M., BARTON, P. & COAST, J. 2017. The Estimation and Inclusion of Presenteeism Costs in Applied Economic Evaluation: A Systematic Review. *Value Health,* 20**,** 496-506.
KIND, P., HARDMAN, G. & MACRAN, G. 1999. *UK population norms for EQ-5D* [Online]. York: Centre for Health Economics. Available: https://www.york.ac.uk/che/pdf/DP172.pdf [Accessed 17.06.2019].
KINSELLA, S. 2008. *Ten lessons for micro-costing in health economics.* [Online]. Available: http://stephenkinsella.net/WordPress/wp-content/uploads/2008/07/tenlessons.pdf [Accessed 07.08.2019].
KNIES, S., SEVERENS, J. L., AMENT, A. J. H. A. & EVERS, S. M. A. A. 2010. The Transferability of Valuing Lost Productivity across Jurisdictions. Differences between National Pharmacoeconomic Guidelines. *Value in Health,* 13**,** 519-527.
KOHL, H. W., CRAIG, C. L., LAMBERT, E. V., INOUE, S., ALKANDARI, J. R., LEETONGIN, G. & KAHLMEIER, S. 2012. The pandemic of physical inactivity: global action for public health. *The Lancet,* 380**,** 294-305.
KOLT, G. S., SCHOFIELD, G. M., KERSE, N., GARRETT, N., SCHLUTER, P. J., ASHTON, T. & PATEL, A. 2009. The Healthy Steps Study: A randomized controlled trial of a pedometer-based Green Prescription for older adults. Trial protocol. *BMC Public Health,* 9**,** 404.
LARSEN, B., GILMER, T., PEKMEZI, D., NAPOLITANO, M. A. & MARCUS, B. H. 2015. Cost effectiveness of a mail-delivered individually tailored physical activity intervention for

Latinas vs. a mailed contact control. *The International Journal of Behavioral Nutrition and Physical Activity,* 12**,** 140.
LEAR, S. A., HU, W., RANGARAJAN, S., GASEVIC, D., LEONG, D., IQBAL, R., CASANOVA, A., SWAMINATHAN, S., ANJANA, R. M., KUMAR, R., ROSENGREN, A., WEI, L., YANG, W., CHUANGSHI, W., HUAXING, L., NAIR, S., DIAZ, R., SWIDON, H., GUPTA, R., MOHAMMADIFARD, N., LOPEZ-JARAMILLO, P., OGUZ, A., ZATONSKA, K., SERON, P., AVEZUM, A., POIRIER, P., TEO, K. & YUSUF, S. 2017. The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study. *The Lancet,* 390**,** 2643-2654.
LEE, I. M., SHIROMA, E. J., LOBELO, F., PUSKA, P., BLAIR, S. N., KATZMARZYK, P. T. & LANCET PHYSICAL ACTIVITY SERIES WORKING, G. 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. 380**,** 219-229.
LEUNG, W., ASHTON, T., KOLT, G. S., SCHOFIELD, G. M., GARRETT, N., KERSE, N. & PATEL, A. 2012. Cost-effectiveness of pedometer-based versus time-based Green Prescriptions: the Healthy Steps Study. *Australian Journal of Primary Health,* 18**,** 204-11.
LEURENT, B., GOMES, M. & CARPENTER, J. R. 2018. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. *Health Economics,* 27**,** 1024-1040.
LIBERATI, A., ALTMAN, D. G., TETZLAFF, J., MULROW, C., GØTZSCHE, P. C., IOANNIDIS, J. P. A., CLARKE, M., DEVEREAUX, P. J., KLEIJNEN, J. & MOHER, D. 2009. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLOS Medicine,* 6**,** e1000100.
LIPUS, A. 2018. *Social Impact Measurement, Part II: a Fit-For-Purpose Mindset Towards Data* [Online]. Available: https://creativeresearchsolutions.com/2018/06/social-impact-measurement-part-ii-a-fit-for-purpose-mindset-towards-data/ [Accessed 06.08.2019 06.08.2019].
LOUDON, K., TREWEEK, S., SULLIVAN, F., DONNAN, P., THORPE, K. E. & ZWARENSTEIN, M. 2015. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ : British Medical Journal,* 350**,** h2147.
LUO, N., JOHNSON, J. & COONS, S. J. 2010. Using instrument-defined health state transitions to estimate minimally important differences for four preference-based health-related quality of life instruments. *Medical Care,* 48**,** 365-71.
MACKENZIE, K., GOYDER, E. & EVES, F. 2015. Acceptability and feasibility of a low-cost, theory-based and co-produced intervention to reduce workplace sitting time in desk-based university employees. *BMC Public Health,* 15**,** 1294.
MADDISON, R., PFAEFFLI, L., WHITTAKER, R., STEWART, R., KERR, A., JIANG, Y., KIRA, G., LEUNG, W., DALLECK, L., CARTER, K. & RAWSTORN, J. 2015. A mobile phone intervention increases physical activity in people with cardiovascular disease: Results from the HEART randomized controlled trial. *European Journal of Preventive Cardiology,* 22**,** 701-9.
MANCA, A., HAWKINS, N. & SCULPHER, M. J. 2005. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Economics,* 14**,** 487-496.
MARMOT, M., ALLEN, J., GOLDBLATT, P., BOYCE, T., MCNEISH, D., GRADY, M., STRELITZ, J., GEDDES, I., FRIEL, S., PORRITT, F., REINERSEN, E., BELL, R. & ALLEN, M. 2010. Fair society, healthy lives: the Marmot Review: strategic review of health inequalities in England post-2010. London: Marmot Review Team,.
MARMOT, M. & ALLEN, J. J. 2014. Social Determinants of Health Equity. *American Journal of Public Health,* 104**,** S517-S519.
MARTEAU, T. M. 2018. Changing minds about changing behaviour. *The Lancet,* 391**,** 116-117.
MATTHEWS, C. E., CHEN, K. Y., FREEDSON, P. S., BUCHOWSKI, M. S., BEECH, B. M., PATE, R. R. & TROIANO, R. P. 2008. Amount of time spent in sedentary behaviors in the United States, 2003-2004. 167**,** 875-881.
MAYER, S. & BEECHAM, J. 2005. *Client Service Receipt Inventory for Adolescents with Chronic Pain- Parent report* [Online]. Available: http://www.dirum.org/instruments/details/100# [Accessed 18.10.2017].
MICHIE, S., ATKINS, L. & WEST, R. 2014. *The Behaviour Change Wheel: A Guide to Developing Interventions.,* London, Silverback Publishing.
MOGYOROSY, Z. & SMITH, P. 2005. The main methdologocial issues in costing health care services. York, UK: Centre for Health Economics.
MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. & THE, P. G. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine,* 6**,** e1000097.

MOORE, G. F., AUDREY, S., BARKER, M., BOND, L., BONELL, C. & HARDEMAN, W. 2015. Process evaluation of complex interventions: Medical Research Council guidance. 350**,** h1258.
MORRIS, A. S., MURPHY, R. C., SHEPHERD, S. O., HEALY, G. N., EDWARDSON, C. L. & GRAVES, L. E. F. 2019. A multi-component intervention to sit less and move more in a contact centre setting: a feasibility study. *BMC Public Health,* 19**,** 292.
MORRIS, S., DEVLIN, N., PARKIN, D. & SPENCER, A. 2012. *Economic Analysis in Healthcare,* New York, UNITED KINGDOM, Wiley.
MURPHY, S. M., EDWARDS, R. T., WILLIAMS, N., RAISANEN, L., MOORE, G., LINCK, P., HOUNSOME, N., DIN, N. U. & MOORE, L. 2012. An evaluation of the effectiveness and cost effectiveness of the National Exercise Referral Scheme in Wales, UK: a randomised controlled trial of a public health policy initiative. *Journal of Epidemiology and Community Health,* 66**,** 745-53.
NATIONAL CAREERS SERVICE. 2019. *Call centre operator* [Online]. Available: https://nationalcareers.service.gov.uk/job-profiles/call-centre-operator [Accessed 31.07.2019].
NHS DIGITAL. 2018. *Prescription Cost Analysis Data* [Online]. [Accessed 06.08.2019 06.08.2019].
NHS IMPROVEMENT. 2018. *Reference costs* [Online]. Available: https://improvement.nhs.uk/resources/reference-costs/#rc1718 [Accessed 13.05.2019].
NHS IMPROVEMENT. 2019. *Patient-level cost reports (PLICS)* [Online]. Available: https://improvement.nhs.uk/resources/patient-level-cost-reports/ [Accessed 16.07.2019].
NICE. 2014a. *National Institute for Health and Care Excellence: Incorporating economic evaluation* [Online]. Available: https://www.nice.org.uk/process/pmg20/chapter/incorporating-economic-evaluation [Accessed 01.11.2017].
NICE. 2014b. *Physical Activity: Exercise Referral Schemes* [Online]. Available: https://www.nice.org.uk/guidance/ph54 [Accessed 16.06.2019].
NICE. 2018. *Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018)* [Online]. Available: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l [Accessed 16.06.2019].
NIXON, R. M., WONDERLING, D. & GRIEVE, R. D. 2010. Non-parametric methods for cost-effectiveness analysis: the central limit theorem and the bootstrap compared. *Health Economics,* 19**,** 316-333.
NOBLE, S. M., HOLLINGWORTH, W. & TILLING, K. 2012. Missing data in trial-based cost-effectiveness analysis: the current state of play. *Health Econ,* 21**,** 187-200.
O'DONOVAN, G., BLAZEVICH, A. J., BOREHAM, C., COOPER, A. R., CRANK, H., EKELUND, U., FOX, K. R., GATELY, P., GILES-CORTI, B., GILL, J. M., HAMER, M., MCDERMOTT, I., MURPHY, M., MUTRIE, N., REILLY, J. J., SAXTON, J. M. & STAMATAKIS, E. 2010. The ABC of Physical Activity for Health: a consensus statement from the British Association of Sport and Exercise Sciences. *J Sports Sci,* 28**,** 573-91.
O'NEILL, J., TABISH, H., WELCH, V., PETTICREW, M., POTTIE, K., CLARKE, M., EVANS, T., PARDO, J. P., WATERS, E. & WHITE, H. 2014. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of clinical epidemiology,* 67**,** 56-64.
O'CONNELL, S. E., JACKSON, B. R., EDWARDSON, C. L., YATES, T., BIDDLE, S. J. H., DAVIES, M. J., DUNSTAN, D., ESLIGER, D., GRAY, L., MILLER, P. & MUNIR, F. 2015. Providing NHS staff with height-adjustable workstations and behaviour change strategies to reduce workplace sitting time: protocol for the Stand More AT (SMArT) Work cluster randomised controlled trial. *BMC Public Health,* 15**,** 1219.
OFFICE FOR EUROPEAN COMMUNITIES 2003. Time use at different stages of life Results from 13 European countries. Luxemburg: Office for Official Publications of the European Communities.
OLDRIDGE, N. B. 2008. Economic burden of physical inactivity: healthcare costs associated with cardiovascular disease. 15**,** 130-139.
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT. 2015. *Promoting Health, Preventing Disease* [Online]. Available: https://www.oecd-ilibrary.org/content/publication/9780335262274-en [Accessed 10.05.2018].
OVER, E. A. B., WENDEL-VOS, G. C. W., VAN DEN BERG, M., REENEN, H. H. H.-V., TARIQ, L., HOOGENVEEN, R. T. & VAN BAAL, P. H. M. 2012. Cost-effectiveness of counseling and pedometer use to increase physical activity in the Netherlands: a modeling study. *Cost Effectiveness and Resource Allocation : C/E,* 10**,** 13-13.

OWEN, L. & FISCHER, A. 2019. The cost-effectiveness of public health interventions examined by the National Institute for Health and Care Excellence from 2005 to 2018. *Public Health,* 169**,** 151-162.
OWEN, L., MORGAN, A., FISCHER, A., ELLIS, S., HOY, A. & KELLY, M. P. 2012. The cost-effectiveness of public health interventions. *Journal of Public Health,* 34**,** 37-45.
OWEN, L., PENNINGTON, B., FISCHER, A. & JEONG, K. 2017. The cost-effectiveness of public health interventions examined by NICE from 2011 to 2016. *Journal of Public Health***,** 1-10.
OWEN, N., HEALY, G. N., MATTHEWS, C. E. & DUNSTAN, D. W. 2010. Too much sitting: the population health science of sedentary behavior. *Exerc Sport Sci Rev,* 38**,** 105-13.
PATTERSON, R., MCNAMARA, E., TAINIO, M., DE SÁ, T. H., SMITH, A. D., SHARP, S. J., EDWARDS, P., WOODCOCK, J., BRAGE, S. & WIJNDAELE, K. 2018. Sedentary behaviour and risk of all-cause, cardiovascular and cancer mortality, and incident type 2 diabetes: a systematic review and dose response meta-analysis. *European Journal of Epidemiology*.
PAVEY, T. G., ANOKYE, N., TAYLOR, A. H., TRUEMAN, P., MOXHAM, T. & FOX, K. R. 2011a. The clinical effectiveness and cost-effectiveness of exercise referral schemes: a systematic review and economic evaluation. 15.
PAVEY, T. G., ANOKYE, N., TAYLOR, A. H., TRUEMAN, P., MOXHAM, T., FOX, K. R., HILLSDON, M., GREEN, C., CAMPBELL, J. L., FOSTER, C., MUTRIE, N., SEARLE, J. & TAYLOR, R. S. 2011b. The clinical effectiveness and cost-effectiveness of exercise referral schemes: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England),* 15**,** i.
PECUNIA. 2019. *ProgrammE in Costing, resource use measurement and outcome valuation for Use in multi-sectoral National and International health economic evaluAtions,* [Online]. Available: https://www.pecunia-project.eu/ [Accessed 25.04.2019].
PETROU, S. & GRAY, A. 2011. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ (Clinical research ed.),* 342**,** d1548-d1548.
PHE. 2018. *Health matters: health economics- making the most of your budget* [Online]. Public Health England,. Available: https://www.gov.uk/government/publications/health-matters-health-economics-making-the-most-of-your-budget/health-matters-health-economics-making-the-most-of-your-budget [Accessed 16.06.2019].
POITRAS, V. J., GRAY, C. E., BORGHESE, M. M., CARSON, V., CHAPUT, J. P., JANSSEN, I., KATZMARZYK, P. T., PATE, R. R., CONNOR GORBER, S., KHO, M. E., SAMPSON, M. & TREMBLAY, M. S. 2016. Systematic review of the relationships between objectively measured physical activity and health indicators in school-aged children and youth. *Appl Physiol Nutr Metab,* 41**,** S197-239.
POPAY, J., ROBERTS, H. M., SOWDEN, A., PETTICREW, M., ARAI, L., RODGERS, M. & BRITTEN, N. 2006. Guidance on the conduct of narrative synthesis in sytematic reviews. Institute for Health Research.
PUBLIC HEALTH ENGLAND. 2019. *Physical activity: applying All Our Health* [Online]. Public Health England. Available: https://www.gov.uk/government/publications/physical-activity-applying-all-our-health/physical-activity-applying-all-our-health [Accessed 05.07.2020].
RABIN, R., OEMAR, M., OPPE, M., JANSSEN, B. & HERDMAN, M. 2011. *EQ-5D-5L User Guide: Basic Information on How to Use the EQ-5D-5L Instrument,* Rotterdam, EuroQol Group.
RAMSEY, S. D., WILLKE, R. J., GLICK, H., REED, S. D., AUGUSTOVSKI, F., JONSSON, B., BRIGGS, A. & SULLIVAN, S. D. 2015. Cost-Effectiveness Analysis Alongside Clinical Trials II—An ISPOR Good Research Practices Task Force Report. *Value in Health,* 18**,** 161-172.
RIDYARD, C. & HUGHES, D. 2012. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. *Value Health,* 15**,** 650-5.
RIDYARD, C. & HUGHES, D. 2015. Review of resource-use measures in UK economic evaluations. *Unit Costs of Health and Social Care.* London, UK: PSSRU.
ROTHWELL, P. M. 2005. External validity of randomised controlled trials: "To whom do the results of this trial apply?". *The Lancet,* 365**,** 82-93.
RUTTER, H., CAVILL, N., BAUMAN, A. & BULL, F. 2018. *Systems approaches to global and national physical activity plans* [Online]. Bulletin of the World Health Organisation,. Available: https://www.who.int/bulletin/volumes/97/2/18-220533/en/ [Accessed 20th August 2019].
RUTTER, H., SAVONA, N., GLONTI, K., BIBBY, J., CUMMINS, S., FINEGOOD, D. T., GREAVES, F., HARPER, L., HAWE, P., MOORE, L., PETTICREW, M., REHFUESS, E., SHIELL, A., THOMAS, J. & WHITE, M. 2017. The need for a complex systems model of evidence for public health. *The Lancet,* 390**,** 2602-2604.

SAMDAL, G. B., EIDE, G. E., BARTH, T., WILLIAMS, G. & MELAND, E. 2017. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. *Int J Behav Nutr Phys Act,* 14**,** 42.
SANDERS, G. D., NEUMANN, P. J., BASU, A. & ET AL. 2016. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *JAMA,* 316**,** 1093-1103.
SCARBOROUGH, P., BHATNAGAR, P., WICKRAMASINGHE, K. K., ALLENDER, S., FOSTER, C. & RAYNER, M. 2011. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *J Public Health (Oxf),* 33**,** 527-35.
SCHOLES, S. 2017. Health Survey for England 2016: Physical Activity in Adults. Health and Social Care Information Centre.
SCHULTZ, A. B., CHEN, C.-Y. & EDINGTON, D. W. 2009. The Cost and Impact of Health Conditions on Presenteeism to Employers. *PharmacoEconomics,* 27**,** 365-378.
SCHULZ, K. F., ALTMAN, D. G. & MOHER, D. 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ,* 340**,** c332.
SCULPHER, M. 2015. Clinical Trials Provide Essential Evidence, but Rarely Offer a Vehicle for Cost-Effectiveness Analysis. *Value in Health,* 18**,** 141-142.
SEDENTARY BEHAVIOUR RESEARCH NETWORK 2012. Letter to the Editor: Standardized use of the terms “sedentary” and “sedentary behaviours”. *Applied Physiology, Nutrition, and Metabolism,* 37**,** 540-542.
SEVICK, M. A., DUNN, A. L., MORROW, M. S., MARCUS, B. H., CHEN, G. J. & BLAIR, S. N. 2000. Cost-effectiveness of lifestyle and structured exercise interventions in sedentary adults11The full text of this article is available via AJPM Online at http://www.elsevier.com/locate/ajpmonline.: Results of project ACTIVE. *American Journal of Preventive Medicine,* 19**,** 1-8.
SHAW, R., FENWICK, E., BAKER, G., MCADAM, C., FITZSIMONS, C. & MUTRIE, N. 2011. 'Pedometers cost buttons': the feasibility of implementing a pedometer based walking programme within the community. *BMC Public Health,* 11**,** 200.
SHRESTHA, N., KUKKONEN-HARJULA, K. T., VERBEEK, J. H., IJAZ, S., HERMANS, V. & BHAUMIK, S. 2016. Workplace interventions for reducing sitting at work. *Cochrane Database Syst Rev,* 3**,** Cd010912.
SIMPSON, A., FLOOD, C., ROWE, J., QUIGLEY, J., HENRY, S., HALL, C., EVANS, R., SHERMAN, P. & BOWERS, L. 2014. Results of a pilot randomised controlled trial to measure the clinical and cost effectiveness of peer support in increasing hope and quality of life in mental health patients discharged from hospital in the UK. *BMC psychiatry,* 14**,** 30-30.
SKIVINGTON, K., MATTHEWS, L., CRAIG, P., SIMPSON, S. & MOORE, L. 2018. Developing and evaluating complex interventions: updating Medical Research Council guidance to take account of new methodological and theoretical approaches. *The Lancet,* 392**,** S2.
SONG, F., EASTWOOD, A. J., GILBODY, S., DULEY, L. & SUTTON, A. J. 2000. Publication and related biases. *Health Technol Assess,* 4**,** 1-115.
SPORT ENGLAND. 2016. *Model for estimating the Outcomes and Values in the Economics of Sport (MOVES v.2.0)* [Online]. Available: https://www.sportengland.org/our-work/health-and-inactivity/what-is-moves/moves-tool/ [Accessed 20th August 2019].
SQUIRES, H., CHILCOTT, J., AKEHURST, R., BURR, J. & KELLY, M. P. 2016. A Framework for Developing the Structure of Public Health Economic Models. *Value in Health,* 19**,** 588-601.
STAMATAKIS, E., EKELUND, U., DING, D., HAMER, M., BAUMAN, A. E. & LEE, I.-M. 2019. Is the time right for quantitative public health guidelines on sitting? A narrative review of sedentary behaviour research paradigms and findings. *British Journal of Sports Medicine,* 53**,** 377-382.
STRATH SCOTT, J., KAMINSKY LEONARD, A., AINSWORTH BARBARA, E., EKELUND, U., FREEDSON PATTY, S., GARY REBECCA, A., RICHARDSON CAROLINE, R., SMITH DEREK, T. & SWARTZ ANN, M. 2013. Guide to the Assessment of Physical Activity: Clinical and Research Applications. *Circulation,* 128**,** 2259-2279.
SVISTAK, M. & PRITCHARD, D. 2018. Economic Evaluation: What is it good for? London, UK: New Philanthropy Capital,.
TELAMA, R., YANG, X., LESKINEN, E., KANKAANPÄÄ, A., HIRVENSALO, M., TAMMELIN, T., VIIKARI, J. S. & RAITAKARI, O. T. 2014. Tracking of physical activity from early childhood through youth into adulthood. *Med Sci Sports Exerc,* 46**,** 955-62.
THOMPSON, S. & WORDSWORTH, S. 2001. An annotated cost questionnaire for completion by patients. HERU Discussion Paper 03/01, University of Aberdeen Health Economics Research Unit, Aberdeen, UK (2001).

THORN, J. C., BROOKES, S. T., RIDYARD, C., RILEY, R., HUGHES, D., WORDSWORTH, S., NOBLE, S. M., THORNTON, G. & HOLLINGWORTH, W. 2018. Core Items for a Standardized Resource Use Measure (ISRUM). *Value in Health,* 21**,** 640-649.

THORN, J. C., COAST, J., COHEN, D., HOLLINGWORTH, W., KNAPP, M., NOBLE, S. M., RIDYARD, C., WORDSWORTH, S. & HUGHES, D. 2013a. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. *Appl Health Econ Health Policy,* 11**,** 155-61.

THORN, J. C., NOBLE, S. M. & HOLLINGWORTH, W. 2013b. Timely and Complete Publication of Economic Evaluations Alongside Randomized Controlled Trials. *PharmacoEconomics,* 31**,** 77-85.

THORPE, K. E., ZWARENSTEIN, M., OXMAN, A. D., TREWEEK, S., FURBERG, C. D., ALTMAN, D. G., TUNIS, S., BERGEL, E., HARVEY, I., MAGID, D. J. & CHALKIDOU, K. 2009. A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *Journal of Clinical Epidemiology,* 62**,** 464-475.

TORBICA, A., TARRICONE, R. & DRUMMOND, M. 2017. Does the approach to economic evaluation in health care depend on culture, values, and institutional context? *The European Journal of Health Economics*.

TORBICA, A., TARRICONE, R. & DRUMMOND, M. 2018. Does the approach to economic evaluation in health care depend on culture, values, and institutional context? *The European Journal of Health Economics,* 19**,** 769-774.

TOWNSEND, N., WICKRAMASINGHE, K., WILLIAMS, J., BHATNAGER, P. & RAYNER, M. 2015. Physical Activity Statistics 2015. Oxford, UK: British Heart Foundation,.

TREMBLAY, M. S., AUBERT, S., BARNES, J. D., SAUNDERS, T. J., CARSON, V., LATIMER-CHEUNG, A. E., CHASTIN, S. F. M., ALTENBURG, T. M., CHINAPAW, M. J. M., ALTENBURG, T. M., AMINIAN, S., ARUNDELL, L., ATKIN, A. J., AUBERT, S., BARNES, J., BARONE GIBBS, B., BASSETT-GUNTER, R., BELANGER, K., BIDDLE, S., BISWAS, A., CARSON, V., CHAPUT, J.-P., CHASTIN, S., CHAU, J., CHINAPAW, M., COLLEY, R., COPPINGER, T., CRAVEN, C., CRISTI-MONTERO, C., DE ASSIS TELES SANTOS, D., DEL POZO CRUZ, B., DEL POZO-CRUZ, J., DEMPSEY, P., DO CARMO SANTOS GONÇALVES, R. F., EKELUND, U., ELLINGSON, L., EZEUGWU, V., FITZSIMONS, C., FLOREZ-PREGONERO, A., FRIEL, C. P., FRÖBERG, A., GIANGREGORIO, L., GODIN, L., GUNNELL, K., HALLOWAY, S., HINKLEY, T., HNATIUK, J., HUSU, P., KADIR, M., KARAGOUNIS, L. G., KOSTER, A., LAKERVELD, J., LAMB, M., LAROUCHE, R., LATIMER-CHEUNG, A., LEBLANC, A. G., LEE, E.-Y., LEE, P., LOPES, L., MANNS, T., MANYANGA, T., MARTIN GINIS, K., MCVEIGH, J., MENEGUCI, J., MOREIRA, C., MURTAGH, E., PATTERSON, F., RODRIGUES PEREIRA DA SILVA, D., PESOLA, A. J., PETERSON, N., PETTITT, C., PILUTTI, L., PINTO PEREIRA, S., POITRAS, V., PRINCE, S., RATHOD, A., RIVIÈRE, F., ROSENKRANZ, S., ROUTHIER, F., SANTOS, R., SAUNDERS, T., SMITH, B., THEOU, O., TOMASONE, J., TREMBLAY, M., TUCKER, P., UMSTATTD MEYER, R., VAN DER PLOEG, H., VILLALOBOS, T., VIREN, T., WALLMANN-SPERLICH, B., WIJNDAELE, K., WONDERGEM, R. & ON BEHALF OF, S. T. C. P. P. 2017. Sedentary Behavior Research Network (SBRN) – Terminology Consensus Project process and outcome. *International Journal of Behavioral Nutrition and Physical Activity,* 14**,** 75.

TRUEMAN, P. & ANOKYE, N. K. 2013. Applying economic evaluation to public health interventions: the case of interventions to promote physical activity. *Journal of Public Health (Oxford, England),* 35**,** 32-39.

TUNIS, S. R., STRYER, D. B. & CLANCY, C. M. 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *Jama,* 290**,** 1624-32.

VAN HOUT, B., JANSSEN, M. F., FENG, Y. S., KOHLMANN, T., BUSSCHBACH, J., GOLICKI, D., LLOYD, A., SCALONE, L., KIND, P. & PICKARD, A. S. 2012. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health,* 15**,** 708-15.

VAN LIER, L. I., BOSMANS, J. E., VAN HOUT, H. P. J., MOKKINK, L. B., VAN DEN HOUT, W. B., DE WIT, G. A., DIRKSEN, C. D., NIES, H. L. G. R., HERTOGH, C. M. P. M. & VAN DER ROEST, H. G. 2017. Consensus-based cross-European recommendations for the identification, measurement and valuation of costs in health economic evaluations: a European Delphi study. *The European Journal of Health Economics*.

VAN MASTRIGT, G. A. P. G., HILIGSMANN, M., ARTS, J. J. C., BROOS, P. H., KLEIJNEN, J., EVERS, S. M. A. A. & MAJOIE, M. H. J. M. 2016. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). *Expert Review of Pharmacoeconomics & Outcomes Research,* 16**,** 689-704.

VIJAY, G. C., WILSON, E. C. F., SUHRCKE, M., HARDEMAN, W. & SUTTON, S. 2016. Are brief interventions to increase physical activity cost-effective? 50**,** 408-417.
VOS, T., CARTER, R., DORAN, C., ANDERSON, I., LOPEZ, A. & WILSON, A. 2007. *Assessing Cost-Effectiveness in the Prevention of Non-Communicable Disease (ACE–Prevention) Project 2005–09* [Online]. Brisbane. Available: https://pdfs.semanticscholar.org/74cd/ce1df56d6fe72fb036bd99e9d56d30ccf32b.pdf [Accessed 10.04.2019].
WANLESS, D. 2004. Securing good for the whole population: Final Report. London: HM Treasury.
WARBURTON, D. E., CHARLESWORTH, S., IVEY, A., NETTLEFOLD, L. & BREDIN, S. S. 2010. A systematic review of the evidence for Canada's Physical Activity Guidelines for Adults. *Int J Behav Nutr Phys Act,* 7**,** 39.
WARBURTON, D. E. R. & BREDIN, S. S. D. 2016. Reflections on Physical Activity and Health: What Should We Recommend? *Canadian Journal of Cardiology,* 32**,** 495-504.
WEATHERLY, H. L. A., COOKSON, R. A. & DRUMMOND, M. 2014. Economic Evaluation of Public Health Interventions: Methodological Challenges. *In:* CULYER, A. J. (ed.) *Encyclopedia of Health Economics.* USA: Elesvier.
WEATHERLY, H. L. A., DRUMMOND, M., CLAXTON, K., COOKSON, R., FERGUSON, B., GODFREY, C., RICE, N., SCULPHER, M. & SOWDEN, A. 2009. Methods for assessing the cost-effectiveness of public health interventions: Key challenges and recommendations. *Health Policy,* 93**,** 85-92.
WELCH, V. A., NORHEIM, O. F., JULL, J., COOKSON, R., SOMMERFELT, H. & TUGWELL, P. 2017. CONSORT-Equity 2017 extension and elaboration for better reporting of health equity in randomised trials. *BMJ,* 359.
WELCH, V. A., TUGWELL, P. & MORRIS, E. B. 2008. The equity-effectiveness loop as a tool for evaluating population health interventions. *Rev Salud Publica (Bogota),* 10 Suppl**,** 83-96.
WHO. 2018a. *World Health Organisation: Noncommunicable diseases* [Online]. Available: http://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [Accessed 10.08.18 10th August 2018,].
WHO. 2018b. *World Health Organisation: Physical Activity: Key Facts* [Online]. Available: http://www.who.int/en/news-room/fact-sheets/detail/physical-activity [Accessed 30.08.2018].
WILLIAMS, C. M., LIN, C.-W. C. & JAN, S. 2012. Economic analysis of physical activity interventions. *British Journal of Sports Medicine,* 46**,** 422-423.
WILLMOTT, M., WOMACK, J., HOLLINGWORTH, W. & CAMPBELL, R. 2015. Making the case for investment in public health: experiences of Directors of Public Health in English local government. *Journal of Public Health,* 38**,** 237-242.
WILLMOTT, M., WOMACK, J., HOLLINGWORTH, W. & CAMPBELL, R. 2016. Making the case for investment in public health: experiences of Directors of Public Health in English local government. *Journal of Public Health,* 38**,** 237-242.
WILMOT, E. G., EDWARDSON, C. L., ACHANA, F. A., DAVIES, M. J., GORELY, T., GRAY, L. J., KHUNTI, K., YATES, T. & BIDDLE, S. J. 2012. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia,* 55**,** 2895-905.
WORLD HEALTH ORGANISATION. 1986. *Ottawa Charter for health promotion* [Online]. World Health Organisation,. Available: http://www.who.int/healthpromotion/conferences/previous/ottawa/en/ [Accessed 21st August 2019].
WORLD HEALTH ORGANISATION 2013. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013–2020. Geneva, Switzerland: World Health Organisation,.
WORLD HEALTH ORGANISATION 2014. Development of the health economic assessment tools (HEAT) for walking and cycling. Copenhagen, Denmark: WHO Regional Office for Europe.
WORLD HEALTH ORGANISATION 2017. WHO methods and data sources for global burden of disease estimates 2000-2015. Geneva: WHO.
WORLD HEALTH ORGANISATION 2018. More Active People For A Healthier World. Geneva, Switzerland: World Health Organisation,.
WU, S., COHEN, D., SHI, Y., PEARSON, M. & STURM, R. 2011. Economic analysis of physical activity interventions. *Am J Prev Med,* 40**,** 149-58.
YOUNG, L., ZIMMERMAN, L., POZEHL, B., BARNASON, S. & WANG, H. 2012. Cost-effectiveness of a symptom management intervention: improving physical activity in older women following coronary artery bypass surgery. *Nurs Econ,* 30**,** 94-103.
ZHAI, L., ZHANG, Y. & ZHANG, D. 2015. Sedentary behaviour and the risk of depression: a meta-analysis. *British Journal of Sports Medicine,* 49**,** 705.

Appendices

Below is a list of appendices presented in this thesis in chronological order:

Chapter 2:

i. Appendix A.1. Search strategy
ii. Appendix A.2. Updated scoping search
iii. Appendix A.3. Data extraction form
iv. Appendix A.4. Quality Assessment
v. Appendix A.5. Supplementary searches

Chapter 3:

i. Appendix B.1. Cost measurement tools

Chapter 4:

i. Appendix C.1. Microcosting exercise
ii. Appendix C.2. Unit cost calculations
iii. Appendix C.3. Participant costs data collection tools
iv. Appendix C.4. Disaggregated costs and consequences

Chapter 5:

i. Appendix D.1. Intervention materials
ii. Appendix D.2. Microcosting exercise
iii. Appendix D.3. Unit cost calculations
iv. Appendix D.4. Disaggregated costs and consequences

Chapter 6:

i. Appendix E.1. Microcosting tool
ii. Appendix E.2. Example of resource use quantification
iii. Appendix E.3. Example of presentation of results

Appendix A.1. Search strategy

Example of search strategy in Medline (Jan 2009- March Week 1 2017)

MEDLINE Ovid Jan Week 1 2009-March Week 1 2017				
Search	Terms	Type	Results	Justification
1	exp exercise/	Controlled vocabulary	153597	• Used in Cochrane Review (likely to be informed by information specialist): http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008366.pub3/full • Exploded as specific terms look relevant
2	exp physical fitness/	Controlled vocabulary	25466	• Used in Cochrane Review(likely to be informed by information specialist): http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008366.pub3/full • Exploded as specific terms looked relevant
3	*Exercise therapy	Controlled vocabulary	22422	• Used in BMJ SR on PA brief interventions: http://bjsm.bmj.com/content/early/2015/10/05/bjsports-2015-094655.full • Focused the search (*) as description of term was not wholly relevant
4	"Physical Education and Training"	Controlled vocabulary	13386	• Used in Cochrane Review(likely to be informed by information specialist): http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008366.pub3/full • Did not explode as no specific terms available
5	Recreation/ OR Dancing/ OR Gardening/ OR Hobbies/ OR exp "Play and Playthings"/ OR exp Sports/ OR exp Relaxation	Controlled vocabulary	192825	• Used in Cochrane Review(likely to be informed by information specialist): http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008366.pub3/full • Exploded where specific terms and definition/or were relevant
6	Physical exertion	Controlled vocabulary	55942	• Used in BMJ SR on PA brief interventions: http://bjsm.bmj.com/content/early/2015/10/05/bjsports-2015-094655.full • Did not explode as no specific terms
7	Life style/ OR healthy lifestyle/ OR sedentary lifestyle	Controlled vocabulary	55760	• Typed in sedentary and lifestyle came up • Did not explode life style as not all specific terms were relevant
8	OR/ 1-7	OR	368620	• Combined controlled vocabulary related to PA and SB
9	(start* or sustain* or maintain* or promot* or uptak* or increas* or	Free text	8669486	• Terms taken from Cochrane Review and BMJ article • Some terms from online thesaurus

	improv* or adher* or encourag* prevent* or reduc* or decreas* or discourag* or chang*) ab,ti.			
10	8 AND 9	AND	206888	• Combined controlled vocabulary terms for PA with free text terms related to increase/ decrease
11	((start* or sustain* or maintain* or promot* or uptak* or increas* or improv* or adher* or encourag* prevent* or reduc* or decreas* or discourag* or chang*) adj3 (inactiv* or activ* or exercis* or fit* or gym* or desk* or station* or sit* or sedentary or stand* or sport* or walk* or lifestyle* or life-style*)).ab,ti.	Free text	670297	• Terms taken from Cochrane Review and BMJ article • Some terms from online thesaurus
12	10 OR 11	OR	811698	
13	Exp Economics/	Controlled vocabulary	548321	• Used in BMJ SR on PA brief interventions: http://bjsm.bmj.com/content/early/2015/10/05/bjsports-2015-094655.full
14	Exp "Costs and Cost Analysis"/	Controlled vocabulary	208732	• Controlled vocabulary • Exploded as done in BMJ article and includes relevant specific terms e.g. Cost Allocation, Cost-Benefit Analysis
15	Exp models, economic/	Controlled vocabulary	12614	• Controlled vocabulary • Exploded as done in BMJ article
16	(cost* or money* or pric* or economic* or budget*) adj2 (effect* or benefit* or utilis* or utilit* or valu* or consequence* or minim* or evaluat* or analys* or apprais* or assess* or model*) ab,ti.	Free text	140074	• Controlled vocabulary • Terms used in BMJ article and Economic Evaluation SR PHI (2016) • Some terms from online thesaurus
17	OR/13-16	OR	625912	• Combined controlled vocabulary and free text terms for economics
18	12 AND 17	AND	10522	• Combined PA and SB terms with economics
19	Add limits: English, 2009- March Week 2 2017, Humans	Limits	4,120	

Appendix A.2. Updated scoping search

A description of the search conducted is described below. Results of the updated searches are presented in the Figure 1 (PRISMA flow diagram). Four studies met the inclusion criteria and are presented in Table 1. Notably, one study presented in Table 1 is an economic evaluation of a SB intervention. The remaining three studies compare PA interventions.

Amendment to limits in Medline (Ovid) Database: 2017- Jan Week 2 2019.

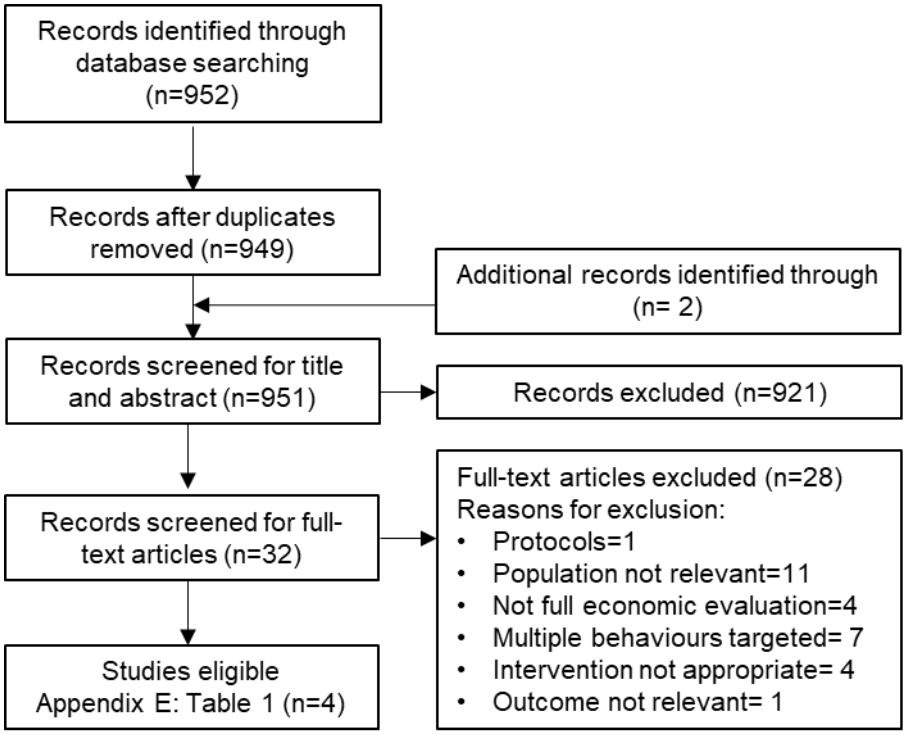


Figure 1. PRISMA flow diagram of study select for updated search (March Week 3 2017-January Week 2 2019)

Database searching

The nine studies retrieved in the original database searching are listed below, all nine came from the Medline Database. This provided the rationale to only rerun the updated search in Medline.

iv. Elley et al. (2011): Medline
v. Boehler et al. (2011): Medline
vi. Pavey et al. (2011): Medline
vii. Anokye et al. (2011): Medline
viii. Edwards et al. (2013): Medline
ix. Young et al. (2012): Medline
x. Iliffe et al. (2014): Medline
xi. Larsen et al. (2015): Medline
xii. Maddison et al. (2015): Medline

Supplementary searching: Key websites

The only websites successful at identifying relevant studies in the original search where the NIHR and NICE evidence websites. These two sites were examined for the updated search (2017-2019). The following study was identified on the NIHR website in the Health Technology Assessment Journal, using the search term “PA”:

Harris T, Kerry S, Victor C, Iliffe S, Ussher M, Fox-Rushby J, *et al.* A pedometer-based walking intervention in 45- to 75-year-olds, with and without practice nurse support: the PACE-UP three-arm cluster RCT. *Health Technol Assess* 2018;22(37)

NICE evidence, searching “PA” (first 10 pages)

Relevant, but already retrieved from NIHR website

Harris T, Kerry S, Victor C, Iliffe S, Ussher M, Fox-Rushby J, *et al.* A pedometer-based walking intervention in 45- to 75-year-olds, with and without practice nurse support: the PACE-UP three-arm cluster RCT. *Health Technol Assess* 2018;22(37)

Excluded due to focus being on population level

PA and the environment: guidance (NG90). Source: National Institute for Health and Care Excellence - NICE - 22 March 2018. This guideline covers how to improve the physical environment to encourage and support PA. The aim is to increase the general population’s PA levels.

Health economic assessment tool (HEAT) for walking and for cycling. Methods and user guide on PA, air pollution, injuries and carbon impact assessments (2017) [PDF]. Source: WHO Regional Office for Europe - WHO Europe - 10 November 2017

Supplementary searching: Key authors

Economic evaluations of targeted SB interventions were searched for via the Google search engine. These included the economic evaluation for key SB interventions: the Stand More at Work intervention in the UK, the Stand More at Work intervention in the US, and the Stand Up Victoria intervention in Australia. Only one economic evaluation was found:

Scandinavian Journal of Work Environment and Health 2018;44(5):503-511 doi:10.5271/sjweh.3740. Economic evaluation of a randomized controlled trial of an intervention to reduce office workers’ sitting time: the "Stand Up Victoria" trial by Gao L, Flego A, Dunstan DW, Winkler EAH, Healy GN, Eakin EG, Willenberg L, Owen N, LaMontagne AD, Lal A, Wiesner GH, Hadgraft NT, Moodie ML

Appendix A.3. Data extraction form

Data extraction form template

1. Study & intervention characteristics
Author & Year
Reviewer
Date Reviewed
Record no. from Endnote
Title
Journal/ Source
Perspective
Vehicle
Intervention
Brief description of intervention
Comparators
Brief description of comparators
Eligible population
Subgoup population
Geographical location
Setting
Sample size
Follow up length of primary data
Duration of effect
Time horizon
Technique
Discount rate
Currency
Price year
Additional info
2. Short term effects in EE
Short term effects identified for EE
Definition of change in PA
Measurement tool for PA
Measurement tool for HRQoL
Measurement tool for other effects in EE (PVO2)
Follow up period of primary data
How are short-term effects expressed?
Additional info
3. Long-term effects in EE
Perspective
Long-term outcomes identified for EE or CCA
Data source for long-term effects
Duration of effect
Time-horizon
How are long-term effects expressed
Additional info
4. Costs
Perspective
Cost categories
Follow up period for primary data
Data source
Additional info

5. Results & Sensitivity
Reporting of results
Base case results
Assessment of uncertainty
Key results reported from uncertainty assessment
Additional info
6. Equity
Equity considerations
Additional info
7. Strengths & Limitations reported by author

Appendix A.4. Quality Assessment

Table 1: Drummond's Checklist Quality Assessment Items 1-5

Study	Q1. Was a well-defined question posted in an answerable form?	Q2 Was a comprehensive description of the competing alternative given? (i.e. can you tell who did what to whom, where, and how often?)	Q3 Was the effectiveness of the programmes or services established?	Q4 Were all the important and relevant costs and consequences for each alternative identified?	Q5 Were costs and consequences measured accurately in appropriate physical units prior to valuation (e.g. hours of nursing time, number of physical visits, lost work-days, gained life-years)?
de Vries et al. 2016	Yes	Yes	Yes	No, societal perspective stated, health sector costs included only	Yes
Pavey et al. 2011	Yes	Yes	Yes	Yes	Can't tell, but reference for secondary source provided
Anokye et al. 2012; Anokye et al. 2014	Yes	Yes	Yes	Yes	Can't tell, but reference for secondary source provided
Larsen et al. 2015	Yes	Yes	Yes	Yes	Yes
Maddison et al. 2015	No, perspective not stated	Yes	Yes	Can't tell, perspective not stated	Can't tell, disaggregated costs not presented
Iliffe et al. 2014	Yes	Yes	Yes	Yes	Yes
Edwards et al. 2013; Murphy et al. 2012	Yes	Yes	Yes	Yes	Yes
Anokye et al. 2011	Yes	Yes	Yes	Yes	Can't tell, but reference for secondary source provided
Elley et al. 2011	Yes	Yes	Yes	Yes	Yes
Shaw et al. 2011	No, perspective not stated	Yes	Yes	Can't tell, perspective not stated	Yes
Over et al. 2012	Yes	Yes	Yes	Yes	Yes
Campbell et al. 2015	Yes	Yes	Yes	Yes	Can't tell, but reference for secondary source provided
Boehler et al. 2011	Yes	Yes	Yes	No, health sector perspective stated, indirect change in healthcare utilisation not included	Yes
Leung et al. 2012	Yes	Yes	Yes	Yes	Yes
Young et al. 2012	Yes	Yes	Yes	No, societal perspective stated, health sector costs included only	Yes

Table 2: Drummond's Checklist Quality Assessment Items 6-10

Study	Q6 Were costs and consequences valued credibly?	Q7 Were costs and consequences adjusted for differential timing?	Q8 Was an incremental analysis of costs and consequences of alternatives performed?	Q9 Was uncertainty in the estimates of costs and consequences adequately characterised?	Q10 Did the presentation and discussion of study results include all issues of concern to the users?	Number of items scored 'No'
de Vries et al. 2016	Yes	Yes	Yes	Yes	Yes	1
Pavey et al. 2011	Yes	Yes	Yes	Yes	Yes	0
Anokye et al. 2012; Anokye et al. 2014	Yes	Yes	Yes	Yes	Yes	0
Larsen et al. 2015	Yes	Yes	Yes	Yes	Yes	0
Maddison et al. 2015	No, price source not stated	Yes	Yes	Yes	Yes	2
Iliffe et al. 2014	Yes	Yes	Yes	Yes	Yes	0
Edwards et al. 2013; Murphy et al. 2012	Yes	Yes	Yes	Yes	Yes	0
Anokye et al. 2011	Yes	Yes	Yes	Yes	Yes	0
Elley et al. 2011	Yes	No, two year follow up	Yes	Yes	Yes	1
Shaw et al. 2011	Yes	Yes	Yes	Yes	Yes	1
Over et al. 2012	Yes	Yes	Yes	Yes	Yes	0
Campbell et al. 2015	Yes	Yes	Yes	Yes	Yes	0
Boehler et al. 2011	Yes	Yes	Yes	Yes	Yes	1
Leung et al. 2012	Yes	Yes	Yes	Yes	Yes	0
Young et al. 2012	Yes	Yes	Yes	Yes	Yes	1

Appendix A.5. Supplementary searches

Table 1. Supplementary searches: grey literature

Websites	**Date searched**	**No. new studies**
NIHR Public Health	02.06.2017	1
NICE (evidence search)	02.06.2017	1
International Society for Pharmoeconomics and Outcomes Research (scientific presentations database)	05.06.2017	0
Centre for Diet and Activity Research (publications database)	05.06.2017	0
National Obesity Observatory (National Archives Gov.UK)	05.06.2017	0
Gov.UK	05.06.2017	0
Fuse.ac.uk	05.06.2017	0
DECIPHer (Development and Evaluation of Complex Interventions for Public Health Improvement)	05.06.2017	0
UK Economic and Social Research Council	05.06.2017	0
Publications from School of Sport, Exercise and Health Sciences at Loughborough University	05.06.2017	0
Centre for Excellence (Northern Ireland)	05.06.2017	0
Centre for Disease Control and Prevention (US gov)	05.06.2017	0
http://www.greylit.org/home	07.06.2017	0
http://www.opengrey.eu/	07.06.2017	0
Reference lists of systematic reviews	**Date searched**	**No. new studies**
Vijay et al., 2016	07.06.2017	2
Wu et al., 2011	07.06.2017	0
Protocol search for completed study	**Date searched**	**No. new studies**
de Vries et al., 2013	07.06.2017	1
Kolt et al., 2009	07.06.2017	1

Appendix B.1. Cost measurement tools

Table 1. Cost items and measurement tools

Study	Leung et al. (2012)	Larsen et al. (2015)	Illiffe et al. (2014)	Edwards et al. (2013)	Elley et al. (2011)	Over et al. (2012)	Isaac et al. (2007) used by Campbell et al. (2015), Payey et al. (2011), Anokye et al. (2011)
Perspective	Public health system and participant	Payer	NHS (also collect participant/private costs)	Multi-agency public sector	Societal	Health care	Societal
Intervention costs							
Data collection method	Not reported	Not reported	Study trial records	Budget breakdown of 13 ERS from Welsh Government; Telephone interviews with NERS programme directors at WG and leisure centre managers to capture additional operating and set up costs	Study trial records	Bottom-up costing of theoretical intervention; Not primary data	Microcosting exercise based on planned intervention sessions
Room hire	Not reported	Not reported	Hall hire	LA annual operating costs= room hire	Not reported	Not reported	Not reported
Equipment	Not reported	Pedometer Computer Printer Scanner Scanner software Software for updating the expert system Printed booklets Scanner paper Binders	Refreshments Mats Resistance bands Instruction booklet	National annual operating costs= printing and stationary LA authority annual operating costs= promotional material	Not reported	Pedometer with electronic diary	Facilities (not described however)
Staff time and salary	Common costs for coordinating programme and telephone counselling excluded GP visits	Trainer salary (with 44% fringe benefits and 10% overheads) Research assistant salary (with 44% fringe benefits and 10% overheads)	Community physiotherapist salary (including preparation, clear-up and travel time)	National annual operating costs= Salary of PA specialist (0.8 WTE) Salary Line Management Grade 7 (0.02 WTE)	Time spend delivering intervention Primary care nurse; Time spent on telephone support	GP assistant approaching patients; GP checking PA level; GP counselling; 3x GP assistant follow up sessions	Exercise trainers

				Salary Executive officers (1.2 WTE) Meeting costs National resources Exercise professionals (36 WTE) Training Travel Joint national and local authority annual operating cost= coordinator salary and on-costs (13 WTE, 50% funded by WG & LA) Local authority= Staff management, attending conferences	by Regional sport trust staff		
Administrative support	Not reported	Postage of intervention material Postage of questionnaire with first class return stamp	Not reported	National annual operating costs: administration;	Not reported	Not reported	Administrative support
Set up costs							
Data collection method	Not reported	Not reported	Not reported	Budget breakdown of 13 ERS from Welsh Government; Telephone interviews with NERS programme directors at WG and leisure centre managers to capture additional operating and set up costs	Not reported	Not reported	Not reported
Staff time and salary	Not reported	Not reported	Not reported	PA specialist (0.2 WTE in yr 1, 0.6 WTE yr 2) Line Management (0.02 WTE yrs 1&2 Grade 7)	Not reported	Not reported	Not reported

				Executive officer (0.2 WTE yr 2) Local authority staff attending meetings			
Meetings	Not reported	Not reported	Not reported	Meetings	Not reported	Not reported	Not reported
Equipment	Not reported	Not reported	Not reported	Resources & printing; IT; Staff clothing; Promotion or advertising; Home working facilities	Not reported	Not reported	Not reported
Translation costs	Not reported	Not reported	Not reported	Translation costs	Not reported	Not reported	Not reported
Training for exercise professionals	Not reported	Not reported	Not reported	Exercise professionals training (Level 3)	Not reported	Not reported	Not reported
Pilot exercise referral project	Not reported	Not reported	Not reported	Costs incurred by WG for the 6 pilot areas in 2006–07 are included as part of NERS development costs (£183,600).	Not reported	Not reported	Not reported
Healthcare utilisation							
Data collection method	Quarterly diaries for 12 months	Not reported	GP records 6 months during intervention 12 months follow up	Knapp M, Beecham J. Costing mental health services: the client service receipt inventory. *Psychol Med* 1990;20:893-908; Baseline, 6 and 12 months	12 months before each assessment	Not reported, as model using secondary data	Case-note review of GP records 12 and 6 months before and after the intervention start date
Primary care contacts	GP visits Physiotherapy visits Nurses Home help Other allied health professionals (acupuncturist, dietician, occupational therapist, podiatrist, social worker, and speech therapist)		Number of contacts with GP, Nurse, Out of hours service, other senior-level practitioners (community matron, specialist nurse, counsellor, pharmacist);, other middle-level practitioners (district nurse,	GP consultation: in surgery; at home, telephone. Practice nurse consultation; Mental health professional	GP visit non-accident related; GP accident-related; GP after hours. Visits to other health providers/ allied health therapy (e.g. physiotherapist, chiropractors, osteopaths, occupational	GP checking PA level and counselling session; GP assistant approaching patient and 3 follow up sessions;	GP visits

			allied health professionals), lower-level practitioners (health-care assistant, support worker, phlebotomist, podiatrist): Contact at practice or Home visits or Telephone calls		therapists, acupuncturists) including after hours.		
Pharmaceuticals prescribed	Not reported	Not reported	Not reported	Primary care prescribing	Not reported	Not reported	Pharmaceuticals prescribed by GP
Secondary care contacts	Admitted to patient as inpatient Hospital specialist Hospital outpatient test/procedure	Not reported	Number of A&E visits for falls; Number of hospital admissions for falls; Number of inpatient nights for falls	Outpatient consultant; Outpatient specialist; Outpatient Physiotherapist; Outpatient other hospital attendances; Day cases; Inpatient hospital days (all causes) A&E attendances Tests	Inpatient hospital admissions; outpatient initial visit; outpatient follow up visit; emergency department visit	Not reported	Hospital admissions (day case and inpatient)
Participant costs							
Data collection methods	Recorded by participants in quarterly diaries for 12 months	Not reported	Diary returns (six during the intervention, 4 in the subsequent 12 month follow up period); Travel costs= At the end of the intervention (6months), participants asked usual method of getting to the class; At post intervention assessment asked what activity they had given up to attend the exercise classes.	Self-completed questionnaire at 12 months: 1 willingness-to-pay (WTP) question for exercise classes;	Self-reported face-to-face questionnaire delivered by a research nurse at baseline, 12 months and 24 months (assessing 12 months before each assessment)	Not reported, as model using secondary data	Self-completed questionnaire 10 weeks follow up assessment

Out of pocket costs	Sports equipment PA fees/ classes	Not reported	Clothes Equipment Gym membership Travel costs to exercise classes	WTP per session	Exercise costs: purchase of exercise or sports shoes, membership fees to exercise groups or gyms, costs of exercise groups or gyms, costs of exercise equipment. Transport/ travel costs to and from the location of exercise or PA Any other additional costs associated with exercise	Not reported	Equipment Childcare Travel cost
Productivity cost	Not reported	Not reported	Time given up to attend the classes	Not reported	Sick day leave Accident-related leave	Not reported	Time working or non-working
Provider (leisure centre)							
Data collection methods	Not reported	Not reported	Not reported	Telephone interviews	Not reported	Not reported	Not reported
Provider productivity cost	Not reported	Not reported	Not reported	Loss in revenue	Not reported	Not reported	Not reported

Health care utilisation questionnaire

It is **very important that you try to answer every question.** If you are unable to remember the exact answers, please try to estimate them as best you can.

Primary Care Services

Question 1: In the last **X months**, which of the following primary care services have you used for any reason (not including hospital appointments recorded in Question 5)?

Please enter '0' in the first row if you have not had any appointments in the last X months.

Service	Tick if YES	Tick if paid for privately	Number of appointments at the clinic/ office/ surgery in the last 6 months?	Number of appointments in your own home in the last 6 months?	Number of appointments over the telephone in the last 6 months?	On average how many minutes did you see/talk to them for each time?
General Practitioner (GP)						
Practice Nurse						
Other allied health professionals						
Physiotherapist						
Chiropractor						
Osteopath						
Occupational therapist						
Acupuncturist						
Specialist Nurse						
District Nurse						
Counsellor						
Pharmacist						
Health-care assistant						
Support worker						
Phlebotomist						
Podiatrist						
Dietician						
Social worker						
Speech therapist						

Mental health professional						
Other (describe) ……..						
Other (describe) ……..						
Other (describe) ……..						

Question 2: Please list **<u>all</u>** medications you have taken in the last X months. Indicate what dose, how often it is taken, and how many weeks you took this medication/ has been taking it. Please list prescription medications only.

Medication name	Dose (mg) (if known)	How many times per day	For how many weeks?

Secondary care services

Question 3: Hospital inpatient care: If you have stayed overnight **in hospital in the last X months for any reason,** please give details below about each stay.

Please enter '0' in the first row if you have not stayed in hospital in the last X months.

Admissions	Speciality of the ward you stayed in (e.g. general ward, surgical ward)	Number of nights for each stay
1st admission		
2nd admission		
3rd admission		
4th admission		
5th admission		
6th admission		

Question 4: Hospital outpatient care: If you have had any **hospital outpatient appointments in the last X months for any reason,** please give details below about each episode.

Please enter '0' in the first row if you haven't had any appointments in the last X months.

Outpatient department/ Consultant speciality (e.g. Rheumatology, Orthopaedic Surgeon, Pain clinic)	Number of appointments in the last X months

Question 5: Accident & Emergency visits: How many times have you visited a hospital accident and emergency department (A&E, casualty) **in the last X months for any reason**.

Number of visits ……..

Participant's Cost questionnaire

Background information

In this questionnaire, we are trying to find out the costs to you for participating in the Exercise Referral Scheme. Unfortunately, we are unable to reimburse these costs. However, your answers are important because they will give people who make decisions about these services an idea of how much the scheme costs you.

Please answer every question. If you are not sure or cannot remember the exact details, please give the best answer you can.

Section A: Participant travel costs

- Question 1:

Over the last 12 weeks how many times have you attended an Exercise Referral Scheme consultation. Please write the number of times in the box below. Put zero if you have not attended an Exercise Referral Scheme consultation over the last 12 weeks for your consultations.

Number of times …….. (/max 4)

If you answered 1 or more to the question above please continue to **Question 2**. Otherwise, skip to **Section B**.

- Question 2:

When you visited the leisure centre for your consultations, how did you **normally** travel? Please circle the number that best describes how you **normally** travelled from your home to the leisure centre for consultations. If you normally used more than one form of transport, please indicate the way you travelled for the **main** (longest in terms of distance) part of your journey.

Walked 1
Cycled.....2
Bus.....3
Train/metro.....4
Taxi.....5

Private car.....6

Motorbike 7

Other (**please specify**) ……..……8

- Question 3:

If you **normally** travelled by public transport (e.g. bus or train) for part or the entire journey, what was the cost of the one-way fare? Please write the cost in the box below. Put **0** if you did not normally travel by public transport at all or if you did not normally pay a fare.

Cost of one-way fare (£)…....._ ……..pence

If you **normally** travelled by taxi for part or the entire journey, what was the cost of the one-way fare? Please write the cost in the box below. Put **0** if you did not normally travel by taxi at all or if you did not normally pay a fare.

Total cost of one-way fare (£)……_……..pence

If you **normally** travelled by private car or motorbike for part or the entire journey how many miles did you travel one-way? Please write the number of miles in the box below. Put **0** if you did not normally travel by private car or motorbike at all.

Number of miles one-way……

If you **normally** travelled by private car or motorbike for part or the entire journey and had to pay tolls or parking fees how much did these amount to? Please write the cost in the box below. Put **0** if you did not normally travel by private car or motorbike at all or did not normally pay tolls or parking fees.

Expenditure on tolls or parking fees (£)…….…….pence

- Question 4:

When you visited the leisure centre, how long did it normally take to travel there from your home? Please write the number of hours and minutes in the box below.

Number of hours……..…..minutes

Section B: Participant Time Costs

Question 5:

When you visited the leisure centre for your consultations, how long did you **normally** spend there? Please write the number of hours and minutes in the box below. Include in your answer the time you **normally** spent waiting and the time you **normally** spent with the Exercise Referral Practitioner.

Number of hours……_.....minutes

- Question 6:

What would you **normally** have been doing as your **main** activity if you had not gone to the Exercise Referral consultations? Please circle the number that best describes what you **normally** would have been doing as your main activity if you had not gone to the consultations.

Housework..... 1
Childcare..... 2

Caring for a relative or friend..... 3

Voluntary work..... 4

Leisure activities..... 5

Attending school or university..... 6

On sick leave..... 7

Seeking work 8

Paid work..... 9

Other (please specify)……..... 10

If you **normally** took time off from paid work (or business activity if self-employed) please continue with **Question 7**. Otherwise, go to **Section C**.

- Question 7:

If you took time off from paid work (or business activity if self-employed) to go to the consultations at the leisure centre approximately how much time did you **normally** take off work (or business activity if self-employed)? Please write the number of hours and minutes in the box below.

Number of hours……_.....minutes

Did you **normally** lose earnings as a result? Please circle the appropriate answer?

Yes……1

No……2

What is your **main** occupation?

……

Section C: Other costs

- Question 8:

In the **last 3 months** have you incurred any other costs because of taking part in the Exercise Referral Scheme (e.g. induction cost, purchasing gym membership, equipment and/ or clothing purchased because of participating in the intervention, paying for sessions)?

Yes 1

No 2

If yes, what were they for and how much did you spend? In the table below please write the purpose of other costs and the amount of money spent.

Purpose	Amount spent
	£……-….p
	£……-….p
	£……-….p
	£……-….p
	£……-….p

Do you have any further comments or any information you would like to add about the cost to you of coming to the leisure centre for the Exercise Referral Scheme?

……
……
……

Interview Schedule_Additional operating and set up costs

Cost effectiveness evaluation

Interview/ Focus group Guide - Additional operating and setting up costs for Intervention Personnel (e.g. centre contact/ organisational gatekeeper, team leaders, movement champions)
Section 1: Additional operating costs
1: Have you incurred any additional operating costs because of running the workplace intervention *(e.g. IT, equipment, attending meetings, promotion or advertising)?*

If yes, in the table below please write the purpose of the additional setting up costs and the amount of time or money spent.

Purpose (Please detail the staff required and/or equipment)	Time spent/ Amount
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p

Do you have any further comments or any information you would like to add about the additional cost of operating the workplace intervention?

2: Is there anything else that you would like to tell us about operating the workplace intervention or this interview?

Section 2: Additional setting up costs
Question 3: Have you incurred any additional setting up costs because of running the workplace intervention *(e.g. training, meetings, resources & printing, room hire, administration, travel)?*

If yes, in the table below please write the purpose of the additional setting up costs and the amount of time or money spent.

Purpose (Please detail the staff required and/or equipment)	Time spent/ Amount
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p
	…….mins/ £……-….p

Do you have any further comments or any information you would like to add about the additional setting up costs of the workplace intervention?

Question 4: Is there anything else that you would like to tell us about setting up the workplace intervention or this interview?

Appendix C.1. Microcosting exercise

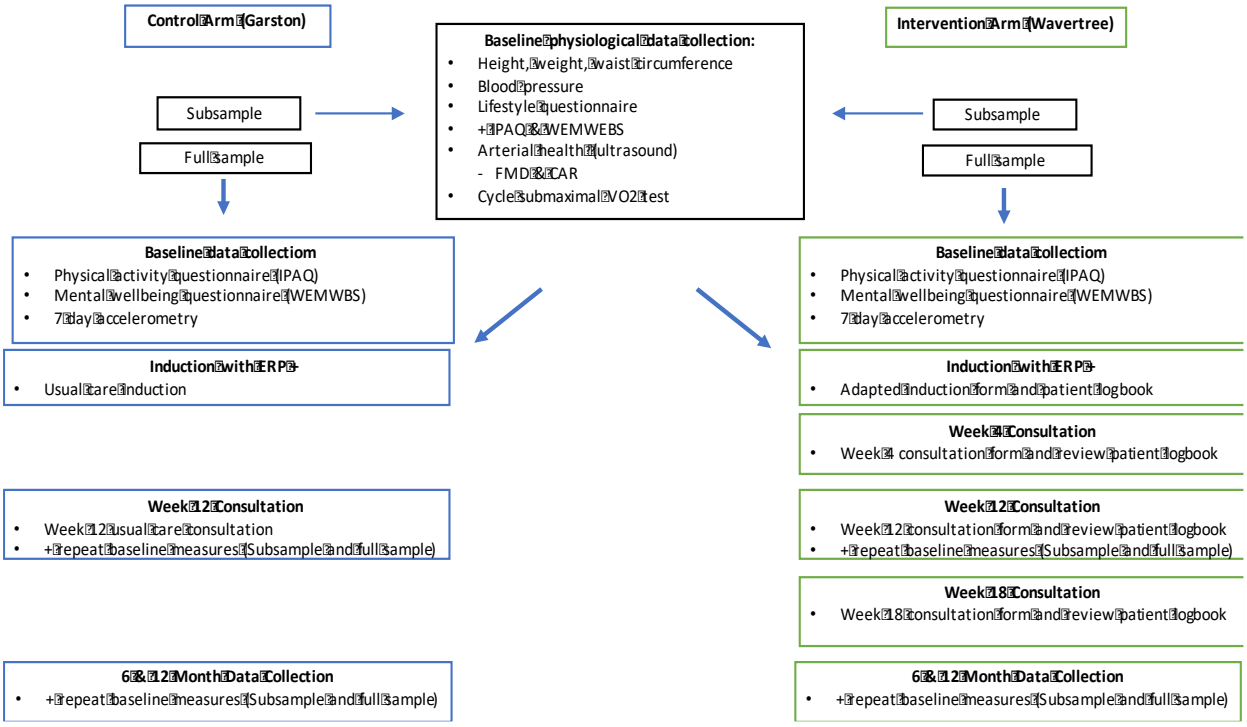


Figure 1. CONSORT Flow Diagram of Usual Care and New Intervention (Co-PARS) from trial protocol

Microcosting tool

The structure of the microcosting exercise tool is based on four types of resources: people, place, programme specific and payments. Definitions for these four concepts have been informed by the ACE-Prevention approach (Vos et al., 2007) and TIDieR framework (Hoffmann et al., 2014).

1. People: Who does what, when and how, and who else is involved
2. Place: Where does it happen and consequently what fixed (capital) equipment (resources) are needed
3. Programme specific: What variable equipment (resources) are needed
4. Payments: What out-of-pocket costs are paid

The tables below, provide examples of the microcosting tables populated in an Excel spreadsheet for the microcosting of the Co-PARS intervention from the public sector perspective. The same struture was used to populate a spreadsheet with data on the usual care intervention.

Table 1.1. People: Who does what, when and how

Activity name	**When (week)**	**1.1. Who (& how)**	1.1. Frequency (number of times)	1.1. Duration (mins)	1.1. Unit cost (£ per minute)	1.1. Total cost (£)
Booking induction	After referral	Receptionist face-to-face, one-to-one	1	5	0.18	0.90
Induction preparation	1-2 days before induction date	ERP face to face, one-to-one	1	6.43	0.43	2.76
Induction/ Call & rebook for no shows	4 weeks after induction booked	ERP face to face, one-to-one	1	60	0.43	25.8
Consultations/ Call backs	Every 4 weeks (week 4, 8, 12, 18)	ERP face to face, one-to-one	4	30	0.43	51.6

Table 1.2. Who else is involved

Activity name	**1.2. Who else: Participant**	1.2. Frequency	1.2. Duration (mins)	1.2. Unit cost (£ £ minute)	1.2. Total cost
Booking induction	Participant	1	5	0	0
Induction preparation	0	0	0	0	0
Induction/ Call & rebook for no shows	Participant	1	60	0	0
Consultations/ Call backs	Participant	4	30	0	0

Table 2.1. Where does it happen

Activity name	**2.1. Where: Capital equipment/ fixed costs**	2.1. Frequency	2.1. Duration (minutes)	2.1. Unit cost (£ per minute)	2.1. Total cost (£)
Booking induction	Leisure Centre reception	1	5	0	0
Induction preparation	Private room in Leisure Centre	1	6.43	0	0
Induction/ Call & rebook for no shows	Private room in Leisure Centre	1	60	0	0
Consultations/ Call backs	Private room in Leisure Centre	4	30	0	0

Table 2.2. What fixed (capital) resources (equipment) are needed

Activity name	**2.2. Where: Capital equipment/ fixed costs**	2.2. Frequency	2.2. Unit cost (£)	2.2.Total cost (£)
Booking induction	IT system at leisure centre	1	0	0
Induction preparation	IT system and telephone at leisure centre	1	0	0
Induction/ Call & rebook for no shows	IT system and telephone at leisure centre	1	0	0
Consultations/ Call backs	IT system and telephone at leisure centre	1	0	0

Table 3.1. What variable equipment (resources) are needed

Activity name	**3.1.Equipment (variable costs)**	3.1. Frequency (pages)	3.1. Unit cost (£ page of printing)	3.1. Total cost (£)
Booking induction	Booking form	1	0.02	0.02
Induction preparation	0	0	0	0
Induction/ Call & rebook for no shows	Participant log book	56	0.07	3.92
Consultations/ Call backs	0	0	0	0

Table 3.2. What other variable equipment (resources) are needed

Activity name	**3.2. Equipment (variable costs)**	3.2. Frequency (months)	3.2. Unit cost (£)	3.2. Total cost (£)
Booking induction	0	0	0	0
Induction preparation	0	0	0	0
Induction/ Call & rebook for no shows	Subsidised membership	3	15	45
Consultations/ Call backs	0	0	0	0

Table 3.3. What other variable equipment (resources) are needed

Activity name	**3.3. Equipment (variable costs)**	3.3. Frequency	3.3. Unit cost (£)	3.3. Total cost (£)
Booking induction	0	0	0	0
Induction preparation	0	0	0	0
Induction/ Call & rebook for no shows	ERP consultation log book	7	0.07	0.49
Consultations/ Call backs	0	0	0	0

Table 3.3. What other variable equipment (resources) are needed

Activity name	**3.4. Equipment (variable costs)**	3.4. Frequency	3.4. Unit cost (£)	3.4. Total cost (£)
Booking induction	0	0	0	0
Induction preparation	0	0	0	0
Induction/ Call & rebook for no shows	Medical questionnaire	2	0.02	0.04
Consultations/ Call backs	0	0	0	0

Table 4.1. What other variable equipment (resources) are needed

Activity name	**4.1. Out of pocket costs (Participants)**	4.1. Frequency	4.1. Unit cost (£)	4.1. Total cost (£)
Booking induction	Induction booking fee	1	7.5	0
Induction preparation	0	0	0	0
Induction/ Call & rebook for no shows	0	0	0	0
Consultations/ Call backs	0	0	0	0

Appendix C.2. Unit Cost calculations

Table 1. Unit cost healthcare utilisation

Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
GP clinic	Cost per 9.22 minutes	Curtis and Burns (2018)	Unit cost used was GP with qualifications but excluding direct care staff. This approach was used by Anokye et al. (2018) Self-reported consultation time by all 3 groups. Mean= 11.9 minutes; Median= 10 minutes; Range= 5-30mins	£34 per 9.22 minute consultation (2018)	No	£34
GP Home visit	Cost per 1 minute= £3.66	Curtis and Burns (2018)	Average travel time of 12 minutes was taken from Curtis (2015) and added to the average clinic consultation time of 9.22 minutes from Curtis and Burns (2018). It was therefore assumed the average time was 21.22 minutes.	£77.66 per 21.22 minute consultation (2018)	No	£77.66
GP telephone call	Cost per 1 minute= £3.60	Curtis and Burns (2018)	Average telephone call of 7.1 minutes was taken from Curtis (2015).	£25.99 per 7.1 minute consultation (2018)	No	£25.99

Practice Nurse clinic[1]	£42 per hour with qualifications	Curtis and Burns (2018)	Unit cost used was Practice Nurse with qualifications. This approach was used by Anokye et al. (2018) £42 per hour equates to £0.70 per minute. Time spent with the practice nurse varied from 5-30minutes, therefore a standard published time was applied, assuming the average consultation time is 15.5 minutes, taken from Curtis (2015).	£10.85 per 15.5 minute consultation (2018)	No	£10.85
Physiotherapist[1,2]	£49.38 per hour with qualifications	Curtis and Burns (2018)	At Band 6, the average salary Band for Physiotherapist (Curtis and Burns, 2018) the cost per working hour is £46 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,410 extra per year for Physiotherapists. £5,410/ working hours per year (1,599)= £3.38 per hour. £46+3.38= £49.38 per hour/ £0.82 per minute (£2018).	£0.82 per minute, consultation length as reported by participant	No	£0.82
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Chiropractor[1,2]	£49.38 per hour with qualifications	Curtis and Burns (2018)	Band 6 was assumed which was similar to the Physiotherapist Band. No qualification costs were provided, and so that reported for physiotherapists were assumed.	£0.82	No	£0.82

Osteopath	n/a	n/a	Osteopathy not typically provided on NHS in England. It is typically sought privately. No participants in this study reported using this service at baseline or 12 weeks, and so a cost was not sought	n/a	n/a	n/a
Occupational Therapist[1]	£47 per hour with qualifications	Curtis and Burns (2018)	Unit cost came from Curtis and Burns (2018) for Community Occupational Therapist including qualifications (£47 per hour/ £0.78 per minute). This approach was used by Anokye et al. (2018)	£0.78	No	£0.78
Acupuncturist	£49.38 per hour with qualifications	Curtis and Burns (2018)	No national unit cost available. Assume it is a Physiotherapist delivering the Acupuncture as described as is typical in NHS in England	£0.82	No	£0.82
Specialist Nurse clinic[2]	£87 per hour	Curtis (2017)	Band 7 Specialist Nurse Cost per hour was £87/ cost per minute was £1,45 (Curtis, 2017)	£1.45 (2017)	Yes. UK's GDP Deflator was applied to inflate the price from 2016/17 to 2018/19 (Multiplier 1.0392). £1.51 per minute	£1.51
District Nurse[2]	£41.73	PSSRU (2012) as cited in Curtis and Burns (2018) as the unit cost for the year 2016/ 2017	District Nurse cost per hour was £41.73 in 2016/ 2017/ cost per minute is £0.70	£0.70 (2017)	Yes. UK's GDP Deflator was applied to inflate the price from 2016/17 to 2018/19 (Multiplier 1.0392). £0.73 per minute	£0.73

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Counsellor[1,2]	£49.38 per hour with qualifications	Curtis and Burns (2018)	In line with the other similar allied Health Professionals Band 6 was used (Curtis and Burns, 2018). The cost per working hour is £46 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,410 extra per year (there is no unit cost for Counsellor or Psychologist qualifications, therefore Physiotherapists unit cost was used: £5,410/ working hours per year (1,599)= £3.38 per hour. £46+3.38= £49.38 per hour/ £0.82 per minute (£2018).	£0.82 per minute, consultation length as reported by participant	No	£0.82
Pharmacist [1,2]	£51.17 per hour with qualifications	Curtis and Burns (2018)	It was assumed the Pharmacist would be a Band 6 (Curtis and Burns, 2018). The cost per working hour is £46 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £8,263 extra per year (1,599 hours per year for Band 6): £5.17 per hour £46+5.17= £51.17 per hour/ £0.85per minute (£2018)	£0.85 per minute, consultation length as reported by participant	No	£0.85
Healthcare assistant clinic[2]	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Healthcare Assistants are Band 2 with an annual	£0.18 per minute	No	£0.18

			pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Support Worker[2]	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Support Workers are Band 2 with an annual pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and	£0.18 per minute	No	£0.18

			unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			
Phlebotomist[2]	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Phlebotomist's are Band 2 with an annual pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.18 per minute	No	£0.18
Podiatrist [1,2]	£49.38 per hour with qualifications	Curtis and Burns (2018)	Band 6 was assumed which was similar to the Chiropractor/ Physiotherapists. No qualification costs were provided, and so that reported for chiropractor/ physiotherapists were assumed.	£0.82	No	£0.82

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Dietician[2]	£48.52 per hour with qualifications	Curtis and Burns (2018)	Unit cost is not broken down by type of visit, therefore the same unit cost was applied for clinic visits, home visits and telephone consultations. At Band 6, the average salary Band for Dietician (Curtis and Burns, 2018) the cost per working hour is £45 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,622 extra per year for Dietician. £5,622/ working hours per year (1,599)= £3.52 per hour. £45+£3.52= £48.52 per hour/ £0.81 per minute (£2018).	£0.81 per minute, consultation length as reported by participant	No	£0.81
Social Worker[1]	£84 per hour with qualifications, client-related work	Curtis and Burns (2018)	Unit cost used was Social Worker with qualifications. This approach was used by Anokye et al. (2018). £84 per hour equates to £1.40 per minute. Time spent with the practice nurse varied from 5-30minutes, therefore a standard published time was applied, assuming the average consultation time is 15.5 minutes, taken from Curtis (2015).	£1.40 per minute, consultation length as reported by participant	No	£1.40

Speech Therapist [1,2]	£48.47 per hour with qualifications	Curtis and Burns (2018)	At Band 6, the average salary Band for Speech Therapist (Curtis and Burns, 2018) the cost per working hour is £45 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,556 extra per year for Speech Therapist. £5,556/ working hours per year (1,599)= £3.47 per hour. £45+£3.47= £48.47 per hour/ £0.81 per minute (£2018).	£0.81 per minute, consultation length as reported by participant	No	£0.81
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Mental Health Professional[2]	£24.24 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Mental Health Nurses. The participants were not asked to specify what type of Mental Health Professional, therefore it was assumed the same Mental Health Professionals delivering the Behavioural Activation interventions would apply (Curtis and Burns, 2018). Wages of Mental Health Nurses are Band 7 (according to Curtis and Burn, 2018) with an annual pay rate from April 2018 of £38,765 for those experienced between 4-5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks	£0.40 per minute	No	£0.40

			(1599 hours) per year. Unit cost= £24.24 per hour or £0.40 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			
Health Trainer[2]	£13.65 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Healthcare Assistants are Band 4 with an annual pay rate from April 2018 of £21,819 for those experienced between 3-4 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £13.65 per hour or £0.28 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.28 per minute	No	£0.28

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Number of hospital admissions for inpatient care (stayed overnight)	£3,894	NHS reference costs (2018)	Hospital admissions for inpatient care, assuming this care is elective (planned), if it was non-elective inpatient care then this would be a lower cost at £1,603 per case. Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups. Type of procedure in secondary care was only recommended as a bolt on module item to collect for studies specifically concerned with extended hospital care because for instance, admissions and re-admissions are prevalent (Thorn et al. 2018) ISRUM Each HRG had an expected bed day, it was assumed participants did not exceed the expected	£3,894	No	£3,894

			bed day and so the exceeded bed day unit cost (£346) was not added on.			
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Number of hospital admissions for day patient care	£742	NHS reference costs (2018)	Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups.	£742	No	£742
Number of hospital outpatient appointments	£125	NHS reference costs (2018)	Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups.	£125	No	£125
Number of visits to A&E	£160	NHS reference costs (2018)	A&E attendance	£160	No	£160

Number of admissions to hospital, after A&E	£1,603	NHS reference costs (2018)	Non-elective inpatient (excluding excess bed days): £1,603 per case. Non-elective means emergency, but can be via GP not just A&E	£1,603	No	£1,603

Footnotes:

1 Unit cost is not broken down by type of visit, therefore the same unit cost was applied for clinic visits, home visits and telephone consultations.

2 Data on the average length of consultation was not provided and so the estimates reported by the participants were used. This was deemed appropriate since only a very small proportion reported this resource.

Table 2. Intervention unit costs

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Receptionist	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Secretary staff are Band 2 with an annual pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to	£0.18 per minute	No	£0.18

			the small magnitude of this cost in terms of the quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			
Exercise referral practitioner	£13.65 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Fitness Instructors are Band 4 with an annual pay rate from April 2018 of £21,819 for those experienced between 3-4 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £13.65 per hour or £0.28 per minute. Published PA Referral Instructor salary for local council, published in 2019 with Level 3 was £13.86-£15 per hour. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.28 per minute	No	£0.28
Grade 6 Researcher	£23.54	Research Institute's pay (2018)	Grade 6 Researcher (mean annual income of £35,929) including salary on-costs (national insurance and superannuation of 14%). University contract of 6 weeks annual leave, 8 bank holidays,	£23.54	No	£23.54

			4 days Christmas closure; 35 hours per week. 1526 hours per year. Hourly unit cost= £23.54			
Printing	£0.07 per A4 page in colour	Research Institute's Printing costs (2018)	Calculation per page: £0.07	£0.07	No	£0.07
Subsidised leisure centre membership	£15 per month	Leisure centre webpage (2019)	Off peak-membership at the leisure centres costed £15 per month: 3 months= £45	£45	No	£45

Table 3. Productivity loss: average weekly earnings and hours worked

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Average UK wage	£14.52 per hour	ONS (2018)	According to the latest data from ONS on average earnings and hours worked in the UK, for 2018 average weekly earnings were £569 (median). Average hours worked per week were 39.2 hours (mean). ONS 2018: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/bulletins/annualsurveyofhoursandearnings/2018. This provides an estimate of £14.52 per hour. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the	£14.52 per hour (2018)	No	£14.52

			quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			

Appendix C.3. Participant costs data collection tools

Participant Travel, Time and Out-of-pocket Costs

Background information

In this questionnaire, we are trying to find out the costs to you for participating in the Exercise Referral Scheme. Unfortunately, we are unable to reimburse these costs. However, your answers are important because they will give people who make decisions about these services an idea of how much the scheme costs you.

Please answer every question. If you are not sure or cannot remember the exact details, please give the best answer you can.

Section A: Participant travel costs

- Question 1:

Over the last 12 weeks how many times have you attended an Exercise Referral Scheme consultation. Please write the number of times in the box below. Put zero if you have not attended an Exercise Referral Scheme consultation over the last 12 weeks for your consultations.

Number of times …….. (/max 4)

If you answered 1 or more to the question above please continue to **Question 2**. Otherwise, skip to **Section B**.

- Question 2:

When you visited the leisure centre for your consultations, how did you **normally** travel? Please circle the number that best describes how you **normally** travelled from your home to the leisure centre for consultations. If you normally used more than one form of transport, please indicate the way you travelled for the **main** (longest in terms of distance) part of your journey.

Walked 1
Cycled.....2
Bus.....3
Train/metro.....4
Taxi.....5

Private car.....6

Motorbike 7

Other (**please specify**) ……..……8

- Question 3:

If you **normally** travelled by public transport (e.g. bus or train) for part or the entire journey, what was the cost of the one-way fare? Please write the cost in the box below. Put **0** if you did not normally travel by public transport at all or if you did not normally pay a fare.

Cost of one-way fare (£)…....._ ……..pence

If you **normally** travelled by taxi for part or the entire journey, what was the cost of the one-way fare? Please write the cost in the box below. Put **0** if you did not normally travel by taxi at all or if you did not normally pay a fare.

Total cost of one-way fare (£)……._……..pence

If you **normally** travelled by private car or motorbike for part or the entire journey how many miles did you travel one-way? Please write the number of miles in the box below. Put **0** if you did not normally travel by private car or motorbike at all.

Number of miles one-way……

If you **normally** travelled by private car or motorbike for part or the entire journey and had to pay tolls or parking fees how much did these amount to? Please write the cost in the box below. Put **0** if you did not normally travel by private car or motorbike at all or did not normally pay tolls or parking fees.

Expenditure on tolls or parking fees (£)……._…….pence

- Question 4:

When you visited the leisure centre, how long did it normally take to travel there from your home? Please write the number of hours and minutes in the box below.

Number of hours……._.....minutes

Section B: Participant Time Costs

Question 5:

When you visited the leisure centre for your consultations, how long did you **normally** spend there? Please write the number of hours and minutes in the box below. Include in your answer the time you **normally** spent waiting and the time you **normally** spent with the Exercise Referral Practitioner.

Number of hours……._.....minutes

- Question 6:

What would you **normally** have been doing as your **main** activity if you had not gone to the Exercise Referral consultations? Please circle the number that best describes what you **normally** would have been doing as your main activity if you had not gone to the consultations.

Housework..... 1
Childcare..... 2

Caring for a relative or friend..... 3

Voluntary work..... 4

Leisure activities..... 5

Attending school or university..... 6

On sick leave..... 7

Seeking work 8

Paid work..... 9

Other (please specify)……..... 10

If you **normally** took time off from paid work (or business activity if self-employed) please continue with **Question 7**. Otherwise, go to **Section C**.

- Question 7:

If you took time off from paid work (or business activity if self-employed) to go to the consultations at the leisure centre approximately how much time did you **normally** take off work (or business activity if self-employed)? Please write the number of hours and minutes in the box below.

Number of hours……_.....minutes

Did you **normally** lose earnings as a result? Please circle the appropriate answer?

Yes……1

No……2

What is your **main** occupation?

……

Section C: Other costs

- Question 8:

In the **last 3 months** have you incurred any other costs because of taking part in the Exercise Referral Scheme (e.g. induction cost, purchasing gym membership, equipment and/ or clothing purchased because of participating in the intervention, paying for sessions)?

Yes 1

No 2

If yes, what were they for and how much did you spend? In the table below please write the purpose of other costs and the amount of money spent.

Purpose	Amount spent
	£……-….p
	£……-….p
	£……-….p
	£……-….p
	£……-….p

Do you have any further comments or any information you would like to add about the cost to you of coming to the leisure centre for the Exercise Referral Scheme?

……
……
……

Willingness to pay Questions

In the three questions below, you will be asked to imagine you were going to take part in a PA programme, and what you would be willing to pay for this.

This information will not impact on any current services you are receiving, but will be used to make recommendations to help inform delivery of PA programmes in future.

Please note there are no minimum or maximum amounts. If you would not be willing to pay anything, please write £0.00.

1. Suppose a one-to-one induction at your local leisure centre involves developing a personalised plan of PA which is tailored to your needs and preferences. The induction is done face-to-face with an Exercise Practitioner and takes 1 hour. Suppose that an induction improves your motivation, confidence and ability to be more physically active. What is the **most** that you would be willing to pay for this induction?

   (£)…..... **:** ……..pence

   Additional comments (e.g. reasons for the amount suggested)

   ……

   ……..

   ……

   ……..

2. Suppose that a swimming session, gym session or exercise class at your local leisure centre improves your motivation, confidence and ability to be more physically active. What is the **most** that you would be willing to pay per swimming session, gym session or exercise class?
Please provide an answer for each type of activity.

Swimming: (£)…..... **:** ……..pence

Gym session: (£)…..... **:** ……..pence

Exercise class: (£)…..... **:** ……..pence

Additional comments (e.g. reasons for the amounts suggested)

……

……..

……

……

3. Suppose you were to take part in 4 one-to-one progress consultations over an 18-week period at your local leisure centre, which involve reviewing your personalised plan of PA (tailored to your needs and preferences). Each consultation is done either face-to-face or over the telephone with an Exercise Practitioner and takes approximately 20 minutes. Suppose that each consultation improves your motivation, confidence and ability to be more physically active. What is the **most** that you would be willing to pay per consultation?
Please provide an answer for both face-to-face and telephone consultations.

Face-to-face: (£)…..... **:** ……..pence

Over the telephone: (£)…..... **:** ……..pence

Additional comments (e.g. reasons for the amounts suggested)

……

……

……

……

Appendix C.4. Disaggregated costs and consequences

Table 1. Incremental QALYs: unadjusted for baseline imbalances

Variable	Co-PARS	Usual Care ERS	Control
	Mean (SD) per participant	Mean (SD) per participant	Mean (SD) per participant
QALYs			
Unadjusted QALYs	0.355 (0.081)	0.364 (0.087)	0.436 (0.064)
Incremental QALYs unadjusted			
Co-PARS vs Usual Care	-0.009 (SE:0.027; 95% CI: -0.062-0.044); this means the Co-PARS group has 0.009 less QALYs compared to usual care		
Co-PARS vs Control	-0.081 (SE: 0.023; 95% CI: -0.127- -0.034); this means the Co-PARS group has 0.081 less QALYs compared to the control group		
Breakdown of Unadjusted QALYs			
Life years	0.5 years	0.5 years	0.5 years
Unadjusted EQ-5D score for all three time points	0.709 (0.160)	0.728 (0.175)	0.871 (0.129)
EQ-5D score at Baseline	0.640 (0.238)	0.724 (0.156)	0.872 (0.135)
EQ-5D score at 12 weeks	0.759 (0.161)	0.751 (0.180)	0.877 (0.121)
EQ-5D score at 6 months	0.729 (0.159)	0.708 (0.242)	0.863 (0.163)

* Adjusted for differences in baseline EQ-5D score using multiple regression

Table 2. Incremental costs: unadjusted for baseline imbalances

Variables	Co-PARS	Usual Care	Control
	Mean (SD) per participant	Mean (SD) per participant	Mean (SD) per participant
Costs			
Unadjusted total costs at 6 Months	£861.80 (£1,073.89)	£538.24 (£1,122.99)	£356.68 (£738.50)
Incremental costs unadjusted			
Co-PARS vs Usual Care	£323.56 (SE:£369.25; 95% CI: £-369.25- £1,016.37); this means the Co-PARS group costs £323.56 more compared to usual care		
Co-PARS vs Control	£505.12 (SE:£291.69; 95% CI: £-66.59-£1,076.83); this means the Co-PARS group costs £505.12 more compared to the control group		

Table 3. Unadjusted ICER result

ICER unadjusted analysis		
Variables	coPARS vs Usual Care	coPARS vs Control
ICER unadjusted for baseline differences	Usual Care ERS dominates Co-PARS. Costing saving result: £35,798 saved per QALY gained	Control group dominates Co-PARS. Cost saving result: £6,255 saved per QALY gained

Table 4. Key activities and intervention operating costs for the Co-PARS and usual care interventions

Key intervention activities	Co-PARS		Usual Care ERS	
	Description of resources consumed	Mean cost	Description of resources consumed	Mean cost
Booking induction	Booking form; Receptionists time (5 mins)	£0.92	Booking form; Receptionist's time (5 mins)	£0.92
1-2 days before induction date	ERP time (45 mins per week for 7 inductions)	£2.76	n/a	£0
Induction delivery/ Call and rebook no shows	ERP time (60 minutes per participant); ERP consultation log book; Participant log book; Medical questionnaire; Free 3 months subsidised membership)	£75.25	ERP time (60 minutes per participant); Personalised Plan; Medical questionnaire; Free 3 months subsidised membership)	£71.33
Consultations/ Call back	ERP time (30 minutes per participant per consultation; 4 consultations in total)	£51.60	n/a	£0
Information about post-scheme options	n/a	£0	Receptionists time (10 minutes) to arrange post-scheme options with participant	£4.30
Total mean cost per participant		£130.53		£76.55

Table 5. Intervention set up costs

	Co-PARS	Total cost	Usual Care		No treatment control	
Design of consultation and participant log books	1 day (7 hours) Grade 6 Researcher	£164.78	n/a	£0	n/a	£0
Roll out meeting	Six meetings to plan the roll out of the intervention at the selected site, average meeting time (2hours). Time of ERP and Grade 6 researcher. Overhead costs included in salaries.	£446.28				
Preparing for training workshop	1 day (7 hours) of preparation by Grade 6 Researcher	£164.78				
Training workshop	2 two day training workshops delivered by the Grade 6 Researcher to ERP	£520.66				
Training one-to-one support	6 hours of one to one support delivered by Grade 6 Researcher to ERP	£223.14				

Reflection on workshop delivery	6 hour reflection on training	£141.24				
Ongoing email and telephone support	Delivered by Grade 6 Researcher to ERP (approximately 2 hours in total)	£74.38				
Total cost	-	£1,735.26				

Table 6. Most frequently visited healthcare professionals in primary care

Resource type	Unit Cost	Co-PARS		Usual care		Control group	
		Mean number of visits (SD); Number of people reporting any use	Total Mean Cost (SD)	Mean number of visits (SD); Number of people reporting any use	Total Mean Cost (SD)	Mean number of visits (SD); Number of people reporting any use	Total Mean Cost (SD)
GP consultations at Baseline	£34 per 9.11 minute consultation	5.60 (5.16); n=24/ 25	£184.78 (£164.60)	2.88 (3.34); n=10/ 16	£93.34 (£107.44)	2.35 (3.41); n=10/14	£72.71 (£98.49)
GP consultations at 6 months		3.00 (2.52); n=20/ 25	£100.08 (£84.36)	2.50 (3.08); n=11/ 16	£60.62 (£71.32)	1.50 (1.51); n=9/14	£49.29 (£49.03)
Practice Nurse consultations at Baseline	£10.85 per 15.5 minute consultation	1.24 (1.69); n=16/ 25	£10.85 (£11.72)	0.94 (1.24); n=8/ 16	£10.17 (£13.41)	2.29 (5.20); n=10/14	£24.80 (£56.37)
Practice Nurse consultations at 6 months		0.75 (0.94); n=14/ 25	£8.14 (£10.24)	0.38 (0.62); n=6/ 16	£4.07 (£6.72)	0.71 (0.99); n=7/14	£7.75 (£10.79)
Physiotherapist visits at Baseline	£0.82 per minute visit, range of average visit was 20-40 minutes	0.64 (1.93); n=7/ 25	£16.40 (£50.27)	2.19 (3.70); n=6/ 16	£107.83 (£270.51)	0.07 (0.27); n=1/14	£1.17 (£4.38)
Physiotherapist visits at 6 months		0.24 (0.60); n=4/ 25	£1.97 (£7.21)	0.31 (1.01); n=2/ 16	£6.15 (324.60)	2.00 (6.42); n=2/14	£4.69 (£17.53)
Counsellor visits at Baseline	£0.82 per minute visit,	0.72 (2.07); n=3/ 25	£24.93 (£86.40)	0.25 (1.00); n=1/ 16	£12.30 (£49.20)	0.86 (3.21); n=1/14	£21.09 (£78.90)

	range of average visit						
Counsellor visits at 6 months	was 50-60 minutes	1.88 (4.16); n=6/ 25	£76.75 (£191.12)	0 (0); n=0/ 16	£0 (£0)	0 (0); n=0/14	£0 (£0)

Table 7. Secondary care healthcare utilisation

Resource type	Unit Cost	Co-PARS		Usual care		Control group	
		Mean resource use (SD); Number of people reporting any use	Total Mean Cost (SD)	Mean resource use (SD); Number of people reporting any use	Total Mean Cost (SD)	Mean resource use (SD); Number of people reporting any use	Total Mean Cost (SD)
Outpatient and Day Case visit at Baseline	£433.50 per visit*	1.96 (2.30); n= 16/ 25	£849.66 (£997.05)	0.81 (0.83); n=10/ 16	£352.22 (£361.62)	1.21 (2.97); n=4/ 14	£526.38 (£1,285.73)
Outpatient and Day Case visit at 6 months		0.84 (1.37); n=9/ 25	£364.14 (£595.96)	0.25 (0.58); n=3/16	£108.38 (£250.28)	0.64 (1.64); n= 3/14	£278.68 (£713.47)
Inpatient admission at Baseline	£3,894 per admission	0.12 (0.33); n=3/ 25	£467.28 (£1,291.49)	0.25 (0.45); n=4/ 16	£973.50 (£1,741.45)	0.15 (0.38); n=2/ 14	£556.29 (£1,414.05)
Inpatient admission at 6 months		0.04 (0.20); n=1/ 25	£155.76 (£778.80)	0.06 (0.25); n=1/ 16	£243.38 (£973.50)	0 (0); n=0/14	£0 (£0)
A&E visits at Baseline	£160 per visit	0.24 (0.44); n=6/ 25	£38.40 (£69.74)	0.19 (0.40); n=3/ 16	£30 (£64.50)	0.07 (0.27); n=1/ 14	£11.43 (£42.76)
A&E visits at 6 months		0.08 (0.28); n=2/ 25	£12.80 (£44.30)	0.06 (0.25); n=1/ 16	£10 (£40)	0 (0); n=0/14	£0 (£0)

*Average cost for outpatient unit cost (£125) and Day case patient unit cost (£742), to be tested in the sensitivity analysis; ** Adjusted for differences in baseline utility using ANCOVA

Table 8. Change in healthcare use for most frequently used services

Primary care			
Resource item	Co-PARS	Usual Care	No treatment control
GP consultations at Baseline	n=24/ 25	n=10/ 16	n=10/14
GP consultations at 6 months	n=20/ 25	n=11/ 16	n=9/14
Practice Nurse at Baseline	n=16/ 25	n=8/ 16	n=10/14
Practice Nurse at 6 months	n=14/ 25	n=6/ 16	n=7/14
Physiotherapist at Baseline	n=7/ 25	n=6/ 16	n=1/14
Physiotherapist at 6 months	n=4/ 25	n=2/ 16	n=2/14
Counsellor at Baseline	n=3/ 25	n=1/ 16	n=1/14
Counsellor at 6 months	n=6/ 25	n=0/ 16	n=0/14
Secondary care			
Resource item	Co-PARS	Usual Care	No treatment control
Outpatient/ Day cases at Baseline; Visiting orthopaedic department (most common department visited)	n= 16/ 25; Orthopaedics: n=2	n=10/ 16; Orthopaedics: n=3	n=4/ 14; Orthopaedics: n=2
Outpatient/ Day cases at 6 months	n=9/ 25; Orthopaedics: n=2	n=3/ 16; Orthopaedics: n=1	n= 3/14; Orthopaedics: n=1
Inpatient admissions Baseline	n=3/ 25	n=4/ 16	n=2/ 14

Inpatient admissions 6 months	n=1/ 25	n=1/ 16	n=0/14
A&E visits Baseline	n=6/ 25	n=3/ 16	n=1/ 14
A&E visits 6 months	n=2/ 25	n=1/ 16	n=0/14

Green: reduction in resource use; Red: increase in resource use

One-way scenario analysis

The aim of the one-way scenario analysis was to assess how sensitive the cost-effectiveness results were to the choice in unit cost for outpatient appointments/ day cases. This was because the healthcare utilisation questionnaire did not ask patients to distinguish between these two activities. In the base case, £433.50 was used as the unit cost, which was an average of the unit costs for outpatients (£125) and day cases (£742). Nevertheless, outpatient appointments were more common than day cases, therefore the one-way scenario analysis assess the impact of using just the outpatient unit cost. As shown in Table 9, the results were consistent with the base case results, whereby Co-PARS group generated was cost-effective (under NICE's threshold) compared to usual care, but not compared to the control group.

Table 9. One-way scenario analysis for outpatient appointment unit costs

Variables	coPARS vs Usual Care	coPARS vs Control
One-way sensitivity analysis Outpatient/Day patient unit cost: ICER unadjusted for baseline differences	Usual Care ERS dominates Co-PARS: £15,661 saved per QALY gained for usual care	No treatment control group dominates Co-PARS: $5,502 saved per QALY gained for usual care
One-way sensitivity analysis Outpatient/Day patient unit cost: ICER adjusted* for baseline utility and costs	£8,439 per QALY	£143,500 per QALY

Table 10. Cost-effectiveness analysis by subgroups within the Co-PARS group

Characteristic	Adjusted ICER*
Most deprived quintile group vs four least deprived quintiles group	Most deprived group dominates (£24 saved per QALY gained for most deprived group)#
Cardiometabolic as main referral reason vs other referral reasons	Cardiometabolic referral reasons dominates (£48,358 saved per QALY gained for cardiometabolic group)#
Males vs Females	£48,286 per QALY (For males Co-PARS is more effective but more expensive)
Aged 55 years and over vs under 55 years	£2,033 per QALY (For those aged 55 and over Co-PARS is more effective but more expensive)#

**Adjusted for baseline healthcare costs and HRQoL score using multiple regression; #Cost-effective/ cost-saving based on NICE's willingness to pay threshold of £30,000 per QALY*

Table 11. Prescribed medications in the last 6 months

Variable	coPARS			Usual Care			No treatment control		
	Baseline	6 months follow up	Change over time	Baseline	6 months follow up	Change over time	Baseline	6 months follow up	Change over time
Answered question	n=16/25 (64%)			n=13/16 (81.25%)			n=10/14 (71.43%)		
At least one medication prescribed	n=14/16 (87.5%)	n=12/16 (75%)	Improved	n=12/13 (92.31%)	n=12/13 (92.31%)	No change	n=9/10 (90%)	n=9/10 (90%)	No change
Any prescribed medications: mean	3.06	2.56	Improved	3.62	2.92	Improved	2.7	2.5	Improved
Any prescribed medications: median	2.5	2	Improved	3	3	No change	2	2.5	Worse
Type: High Blood Pressure	n=5/16 (31.25%)	n=5/16 (31.25%)	No change	n=6/13 (46.15%)	n=6/13 (46.15%)	No change	n=5/10 (50%)	n=5/10 (50%)	No change
Type: High cholesterol	n=4/16 (25%)	n=3/16 (18.75%)	Improved	n=4/13 (30.77%)	n=4/13 (30.77%)	No change	n=2/10 (20%)	n=2/10 (20%)	No change
Type: Antidepressants	n= 3/16 (18.75%)	n=2/16 (12.5%)	Improved	n=3/13 (23.08%)	n=3/13 (23.08%)	No change	n=1/10 (10%)	n=2/10 (20%)	Worse
Type: T2D	n=1/16 (6.25%)	n=1/16 (6.25%)	No change	n=2/13 (15.38%)	n=1/13 (7.69%)	No change	0	0	No change
Type: Moderate to strong painkillers	n=1/16 (6.25%)	0	Improved	n=2/13 (15.38%)	n=2/13 (15.38%)	No change	n=1/10 (10%)	n=2/10 (20%)	Worse
Type: Angina or high blood pressure	n=2/16 (12.5%)	n=2/16 (11.77%)	No change	n=1/13 (7.69%)	n=1/13 (7.69%)	No change	0	0	No change

Participant costs

Table 12. Participant costs

Time (12 weeks)			
Variable	Co-PARS	Usual Care	No intervention control
Number of one-to-one consultations with ERP	Mean: 2.21 Median: 2 Range: 1-4 Answered question correctly: n=19/25	Mean: 1.4 Median: 1 Range: 1-3 Answered question correctly: 10/16	n/a
Travel time one-way from home to leisure centre (two-way/ return)	Mean:19mins (38mins) Median:15mins (30mins) Range:5-60mins (10-120mins) Answered question correctly: 19/25	Mean: 13mins (26mins) Median: 10mins (20mins) Range: 3-60mins (5-120mins) Answered question correctly: 13/16	
Consultation time per visit	Mean: 53mins Median: 40mins Range: 30-120mins Answered question correctly: 19/25	Mean: 62mins Median: 60mins Range: 15-150mins Answered question correctly: 14/16	
Activity displaced due to taking part	Answered question correctly (selected one activity): 15/25 More than one activity named: 7/25 Most common: Leisure Time only n=5/15	Answered question correctly (selected one activity): 9/16 More than one activity named: 5/16 Most common: Housework only n=5/16	

Lost time in work due to attending in consultations	Yes: n=3/15 Loss of earnings due to attending consultations: n=1/3 Number of hours of paid work missed: 1 hour x1 time (n=1/1)	Yes: n=2/9 Loss of earnings due to attending consultations: n=0/2 Number of hours of paid work missed: n/a	
Travel distance (12 weeks)			
Variable	Co-PARS	Usual Care	No intervention control
Most common mode of travel to the leisure centre	Private car n=11/20	Private car n=9/11	n/a
Private car miles travelled	Mean: 2.6 miles Median: 2 miles Range: 1-8 miles Answered question correctly: n=10/11	Mean: 2.4 miles Median: 2.5 miles Range: 1-3 miles Answered question correctly: n=8/9	
Out of pocket costs (12 weeks)			
Variable	Co-PARS	Usual Care	No intervention control
Incurred a cost (excluding induction fee)	Answered question correctly: n=23/25 Incurred a cost: n=13/23	Answered question correctly: n=14/16 Incurred a cost: n=6/14	Answered the question correctly: n=12/14 Incurred a cost: n=7/12
Gym membership cost*	Range: £7-£30 n=7/13	Range: £20-£163 n=2/6	Range: £42-£63 n=5/7
Clothing/ footwear item	Range: £14-£54 n=4/13	Range: £55-90 n=2/6	Range: £50-160 n=2/7

Equipment	Range: £6-£199 (fitbit watch) n=3/13	n=0/6	n=0/7
Class/ Swimming sessions*	Range: £4-£30 n=6/13	Range: £1-£12 n=3/6	Range: £5.50-£72 n=6/7
Outdoor activities	n=0/13	n=0/6	Range: £30-£240 (2xIronman+open water swimming) n=2/7
Personal trainer	£30 per hour x2 n=1/13	n=0/6	0/7
Out-of-pocket costs (6 Months)			
Variable	Co-PARS	Usual Care	No intervention control
Incurred a cost (excluding induction fee)	n=17/25	n=12/15	n=8/14
Gym membership cost*	Range: £15-£150 n=9/17	Range: £15-£150 n=6/12	Range: £20-£56 n=5/8
Clothing/ footwear item	Range: £6-£65 n=6/17	Range: £35-£100 n=2/12	Range: £50-150 n=4/8
Equipment	Price: £22 (bike tyres) n=1/17	n=0/12	n=0/8
Class/ Swimming sessions*	Range: £1-£82 n=8/17	Range: £1-£30 n=5/12	Range: £16.50-£240 n=3/8

Outdoor activities (excluding travel/ accommodation)	n=0/17	n=0/12	Range: £50-£150 (1/2 Ironman) n=2/8
Personal trainer	n=0/17	n=0/12	n=0/8
HCU (Baseline and 6 Months)			
Variable	Co-PARS	Usual Care	No intervention control
Private HCU at Baseline	• 60mins with Acupuncturist (2 at clinic)= £98.40 (n=1) • 30mins with Chiropractor (3 at clinic)= £73.80 (n=1) • 10mins with Podiatrist (4 at clinic)= £65.60 (n=1) • 30mins with Acupuncturist (1 at clinic)= £24.60 (n=1)		• 60mins Sports Massage (6 at clinic)= £270 (n=1)
Private HCU at 6 months	• 60mins with Counsellor (6 at clinic)= £295.20 (n=1) • 45mins with Acupuncturist (6 at clinic)= £221.40 (n=1) • 15mins with Podiatrist (2 at clinic)= £24.60 (n=1)	• GP consultations (10 at surgery; 1 telephone)= £365.99 (n=1)	• 60mins Physiotherapy (24 at clinic)= £393.60 (n=1) • 60mins Sports Massage (1 at clinic; 4 at home)= £225 (n=1) • 45mins with Chiropractor (1 at clinic)= £36.90 (n=1)

*unclear whether referring to per item or aggregate total; **unclear if per same leisure centre as intervention and also if represents per month or past 3 month

Willingness to pay analysis

By large, participants in the control group were willing-to-pay twice as much (£19.64) than the coPARS (£8.23) and usual care (£9.25) group for a hypothetical induction. Similarly, the control group were willing-to-pay twice as much (£13.23) than the Co-PARS (£4.64) and usual care (£6.07) group for a face-to-face consultation. Co-PARS participants appeared to value the swimming sessions slightly more than the gym sessions and exercises classes.

Table 13. Participants' willingness to pay

Willingness to pay (reported at 6 months)			
Variable	coPARS	Usual Care	No intervention control
Willingness to pay per induction	Mean: £8.23 Median: £8.50 Range: £0- £20 Answered question correctly: n=22/25 Most deprived quintile: n=11/22 (50%)	Mean: £9.25 Median: £8.75 Range: £0- £20 Answered question correctly: n=14/16 Most deprived quintile: n=3/14 (21.42%)	Mean: £19.64 Median: £20 Range: £0- £40 Answered question correctly: n=14/14 Most deprived quintile: n=4/14 (21.42%)
Willingness to pay per face-to-face progress consultation	Mean: £4.64 Median: £3 Range: £0- £20 Answered question correctly: n=22/25 Most deprived quintile: n=11/22 (50%)	Mean: £6.07 Median: £5 Range: £0- £20 Answered question correctly: n=15/16 Most deprived quintile: n=3/15 (20%)	Mean: £13.23 Median: £10 Range: £0- £50 Answered question correctly: n=13/14 Most deprived quintile: n=4/13 (30.77%)
Willingness to pay per telephone progress consultation	Mean: £0.64 Median: £0 Range: £0- £5 Answered question correctly: n=19/25	Mean: £1.60 Median: £1 Range: £0- £8	Mean: £6.50 Median: £3.50 Range: £0- £30

	Most deprived quintile: n=10/19 (52.63%)	Answered question correctly: n=15/16 Most deprived quintile: n=3/15 (20%)	Answered question correctly: n=12/14 Most deprived quintile: n=4/12 (33.33%)
Willingness to pay per swimming session	Mean: £3.24 Median: £2.50 Range: £0- £20 Answered question correctly: n=22/25 Most deprived quintile: n=11/22 (50%)	Mean: £2.70 Median: £3 Range: £0- £5 Answered question correctly: n=15/16 Most deprived quintile: n=3/15 (20%)	Mean: £3.93 Median: £5 Range: £0- £10 Answered question correctly: n=14/14 Most deprived quintile: n=4/14 (21.42%)
Willingness to pay per gym session	Mean: £2.70 Median: £2 Range: £0- £10 Answered question correctly: n=22/25 Most deprived quintile: n=11/22 (50%)	Mean: £3.63 Median: £3 Range: £1- £10 Answered question correctly: n=15/16 Most deprived quintile: n=3/15 (20%)	Mean: £6.79 Median: £5 Range: £0- £20 Answered question correctly: n=14/14 Most deprived quintile: n=4/14 (21.42%)
Willingness to pay per exercise class	Mean: £2.50 Median: £2 Range: £0- £5 Answered question correctly: n=21/25 Most deprived quintile: n=10/21 (48.57%)	Mean: £3.10 Median: £3 Range: £0- £5 Answered question correctly: n=15/16 Most deprived quintile: n=3/15 (20%)	Mean: £6.61 Median: £5 Range: £0- £20 Answered question correctly: n=14/14 Most deprived quintile: n=4/14 (21.42%)

Appendix D.1. Intervention materials

LIVERPOOL JOHN MOORES UNIVERSITY

CALL AGENT SCREENING FORM

Agent Name: _

Email: _

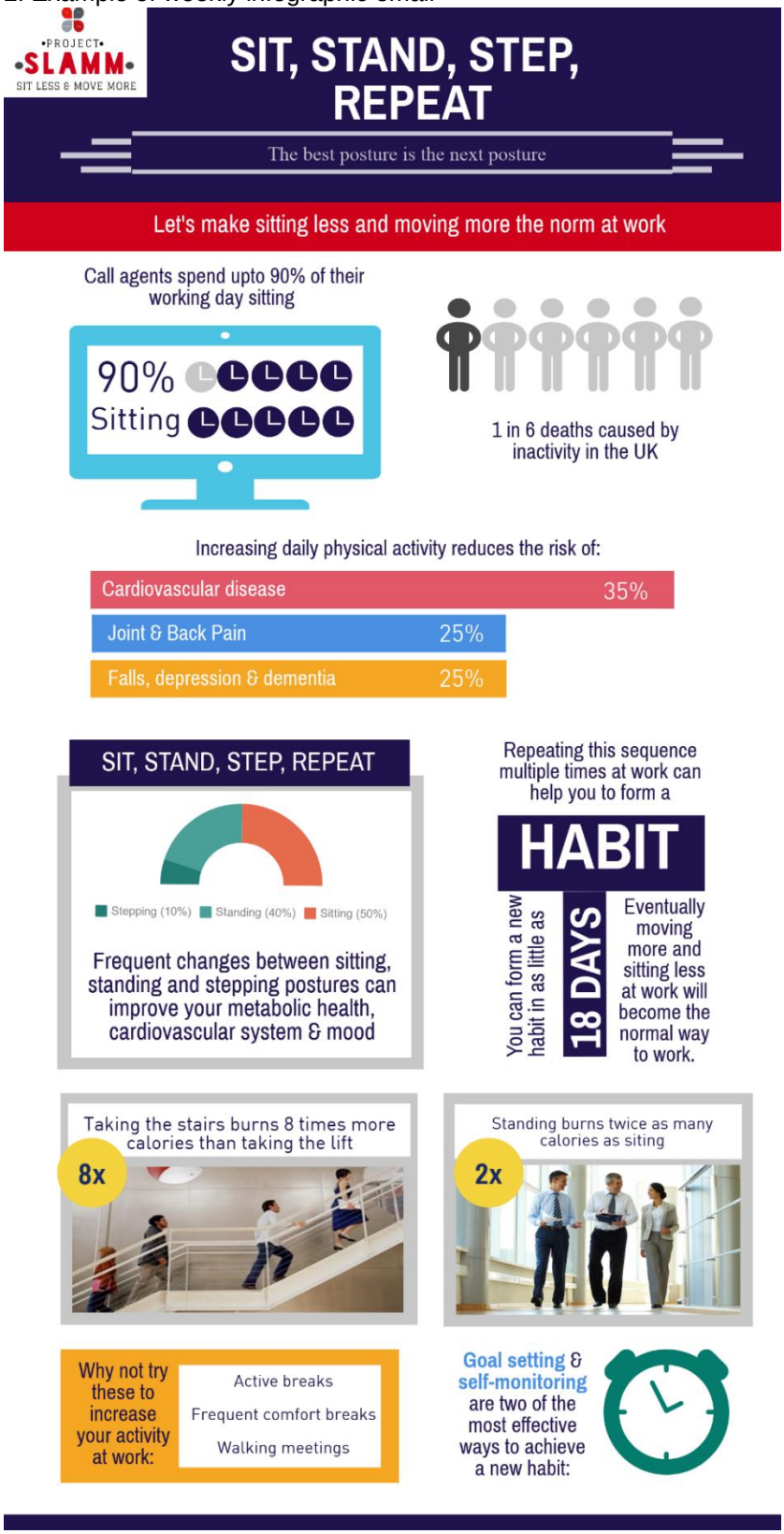
Hours worked per week: _

Contract type: Agency/ Permanent

Team leader: _

Criteria	**Met (Y/N)**	**Comments**
Full time member of staff ≥0.6 full time or part time equivalent worker *(22.5h min)*		
Call agent job role		
Based onsite (Kirkby) throughout the trial period (*July 2018-March 2019*)		
Access to a work telephone and desktop computer with internet		
Aged ≥18 years		
Ambulatory- Able to walk without aid		
No health problems that would impact ability to stand for 10 minutes at a time		
No planned absence >3 weeks during first 3 months of the trial [*July-October 2018*]		
No planned relocation to another workplace/site during the first 3 months of the intervention [*July-October 2018*]		
Not pregnant		

2. Example of weekly infographic email



3. Example of daily log book

Appendix D.2. Microcosting exercise

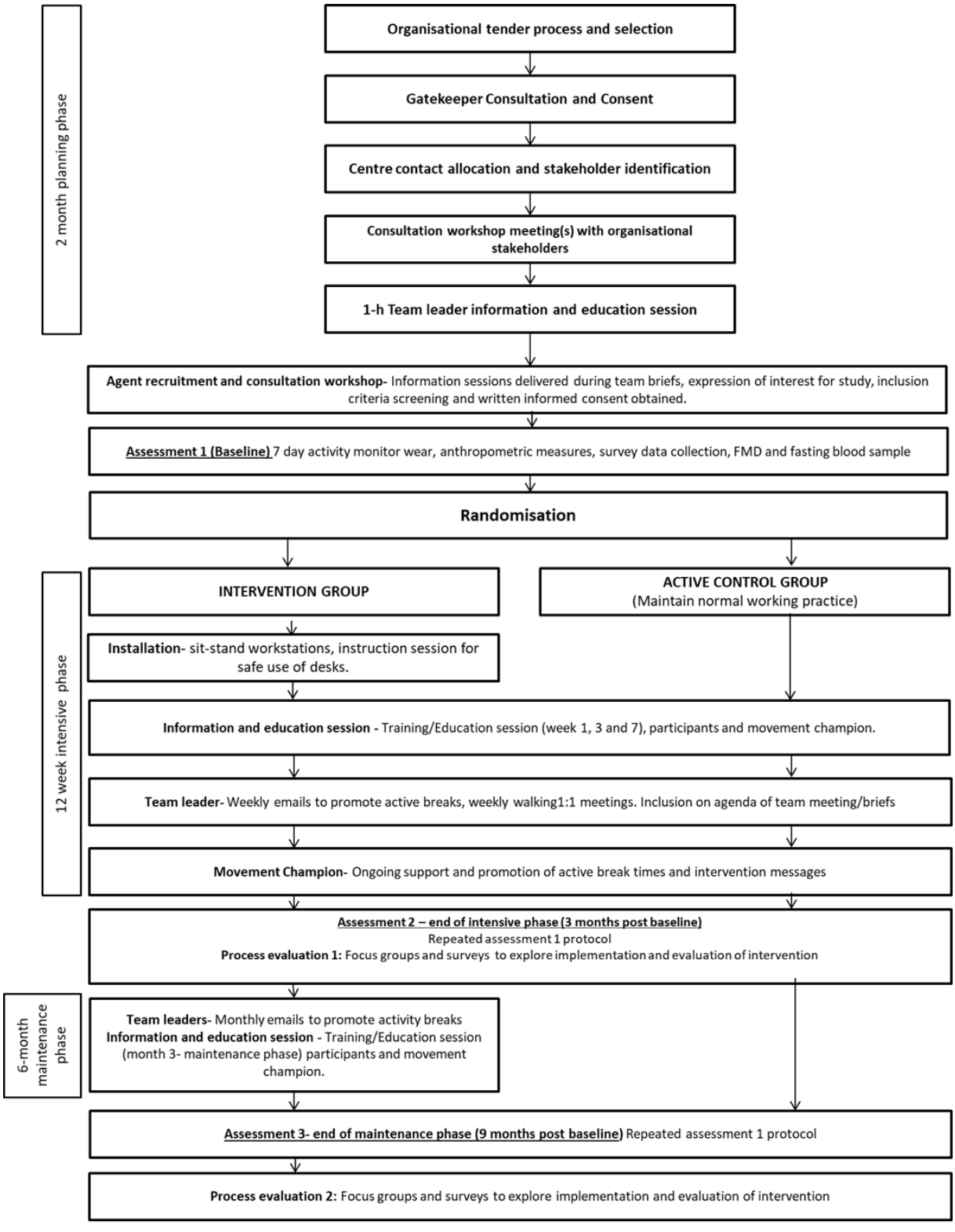


Figure 1. CONSORT Flow Diagram of SLaMM+ intervention and SLaMM (active control)

Weekly electronic log

In Table 1 we are trying to find out about the additional costs to Serco, for setting up the Sit Less and Move More (SLaMM) Project. Please complete the basic information section followed by Table 1. For the table please write the purpose of the additional setting up costs and the amount of time or money you spent. Please refer to the examples in grey for guidance. If you are not sure or cannot remember the exact details, please give the best answer you can. If you have a problem in completing the table, please contact the Researcher Maddy Cochrane from Liverpool John Moores University: m.a.cochrane@2016.ljmu.ac.uk

Question 1: Staff Member Job Title=

Question 2: Staff Salary Band (Optional)=

Table 1. Additional setting up costs

Purpose (Please detail the task type and/or equipment)	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**
Example of time: scheduling Call Agent Briefs	*7 hours*	*5 hours one day and 2 hours on one day the week after*
Example of equipment: printing off posters from Serco Printer to advertise the Sit Less and Move More Project	*50p*	*5 sheets, 10p per sheet*
Example of time: Email contact with other Serco staff about organising Team Leader Briefs	*5 mins*	
Example of time: Meetings about organising Team Leader Briefs	*None*	
Team Leader Briefs	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**

Scheduling **Team Leader Briefs** into Excel Spreadsheets		
Face-to-face meetings about organising **Team Leader Briefs**		
Email contact about organising **Team Leader Briefs**		
Telephone contact about organising **Team Leader Briefs**		
Call Agent Briefs	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**
Scheduling **Call Agent Briefs** into Excel Spreadsheets		
Face-to-face meetings about organising **Call Agent Briefs**		
Email contact organising **Call Agent Briefs**		
Telephone contact about organising **Call Agent Briefs**		
Call Agent Health Checks	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**
Scheduling **Call Agent Health Checks** into Excel Spreadsheets		
Face-to-face meetings about organising **Call Agent Health Checks**		
Email contact about organising **Call Agent Health Checks**		
Telephone contact about organising **Call Agent Health Checks**		

The 1st Call Agent Education and Training Session	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**
Scheduling **the 1st Call Agent Education and Training Session** into Excel Spreadsheets		
Face-to-face meetings about organising **the 1st Call Agent Education and Training Session**		
Email contact about organising **the 1st Call Agent Education and Training Session**		
Telephone about organising **the 1st Call Agent Education and Training Session**		
Other tasks/ materials not detailed above	**Time spent** (mins/ hours) **/ Amount if equipment** (£……-….p)	**Additional information (if possible to add detail)**

Microcosting spreadsheet tool

The structure of the microcosting exercise tool is based on four types of resources: people, place, programme specific and payments. Definitions for these four concepts have been informed by the ACE-Prevention approach (Vos et al., 2007) and TIDieR framework (Hoffmann et al., 2014).

1. People: Who does what, when and how, and who else is involved
2. Place: Where does it happen and consequently what fixed (capital) equipment (resources) are needed
3. Programme specific: What variable equipment (resources) are needed
4. Payments: What out-of-pocket costs are paid

The tables below, provide examples of the microcosting tables populated in an Excel spreadsheet for the microcosting of the SLaMM+ intervention from the public sector and private sector employer's perspective. The same spreadsheet was used to estimate the SLaMM (active control) costs with the deduction of the heigh-adjustable desk costs.

Table 1.1. People: Who does what, when and how

Activity name	**When (week)**	**1.1. Who (& how)**	1.1. Frequency (number of times)	1.1. Duration (hours in decimals format)	1.1. Unit cost (£hour)	1.1. Total cost (£)
Height-adjustable desk	July-October 2017	Agent	20	0	0	0
Weekly emails	July-October 2018	Grade 6 researcher via email	12	0.0833	23.54	23.53
Organisation of Education & Training sessions	At least 1 week before Week 1, 3, 10	Grade 6 researcher via email/ telephone	3	0.25	23.54	17.66
Education & training sessions	Week 1, 3, 10	Grade 6 researcher face to face	12	0.75	23.54	211.86

Table 1.2. Who else is involved

Activity name	**1.2. Who else: Participant**	1.2. Frequency	1.2. Duration (hours in decimals format)	1.2. Unit cost (£per hour)	1.2. Total cost
Height-adjustable desk	0	0	0	0	0
Weekly emails	Team Manager-centre contact	12	0.0833	26.31	26.30
Organisation of Education & Training sessions	Resource planner	3	1.65	19.6	97.02
Education & training sessions	Call agent	120 (40 agents x 3 sessions)	0.5	7.87	472.2

Table 2.1. Where does it happen

Activity name	**2.1. Where: Capital equipment/ fixed costs**	2.1. Frequency	2.1. Duration (hours in decimals format)	2.1. Unit cost (£ per hour)	2.1. Total cost (£)
Height-adjustable desk	Office floor	0	0	0	0
Weekly emails	Via email	12	0	0	0
Organisation of Education & Training sessions	Email/ Telephone	6	0	0	0
Education & training sessions	Private room in contact centre	1	1	0	0

Table 2.2. What fixed (capital) resources (equipment) are needed

Activity name	**2.2. Where: Capital equipment/ fixed costs**	2.2. Frequency	2.2. Duration	2.2. Unit cost (£)	2.2.Total cost (£)
Height-adjustable desk	0	0	0	0	0
Weekly emails	Email	12	0	0	0
Organisation of Education & Training sessions	Email/ telephone	2	0	0	0
Education & training sessions	Contact centre- Travel for Researchers (1 car)	1	14.8	0.15	2.22

Table 3.1. What variable equipment (resources) are needed

Activity name	**3.1.Equipment (variable costs)**	3.1. Frequency (pages)	3.1. Unit cost (£ page of printing)	3.1. Total cost (£)
Height-adjustable desk	Posturite height adjustable desk	19	14.73	279.87
Weekly emails	IT system	1	0	0
Organisation of Education & Training sessions	IT system	3	0	0
Education & training sessions	IT system	3	0.75	0

Table 3.2. What other variable equipment (resources) are needed

Activity name	**3.2. Equipment (variable costs)**	3.2. Frequency (months)	3.2. Unit cost (£)	3.2. Total cost (£)
Height-adjustable desk	Printing of instructions A4	19	0.07	1.33
Weekly emails				0
Organisation of Education & Training sessions				0
Education & training sessions	Timer	40	2.38	95.2

Table 3.3. What other variable equipment (resources) are needed

Activity name	**3.3. Equipment (variable costs)**	3.3. Frequency	3.3. Unit cost (£)	3.3. Total cost (£)
Height-adjustable desk	Lamination cost A4	19	0.18	3.42
Weekly emails				0
Organisation of Education & Training sessions				0
Education & training sessions	Printing log book A7 (8 weeks per page; 2 pages for 12 weeks)	38	0.07	2.66

Table 4.1. What other out of pocket costs are paid

Activity name	**4.1. Out of pocket costs (Participants)**	4.1. Frequency	4.1. Unit cost (£)	4.1. Total cost (£)
Height-adjustable desk	0	0	0	0
Weekly emails	0	0	0	0
Organisation of Education & Training sessions	0	0	0	0
Education & training sessions	0	0	0	0

Appendix D.3. Unit cost calculations

Table 1. Unit cost healthcare utilisation

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
GP clinic	Cost per 9.22 minutes	Curtis and Burns (2018)	Unit cost used was GP with qualifications but excluding direct care staff. This approach was used by Anokye et al. (2018) Self-reported consultation time by all 3 groups. Mean= 11.9 minutes; Median= 10 minutes; Range= 5-30mins	£34 per 9.22 minute consultation (2018)	No	£34
GP Home visit	Cost per 1 minute= £3.66	Curtis and Burns (2018)	Average travel time of 12 minutes was taken from Curtis (2015) and added to the average clinic consultation time of 9.22 minutes from Curtis and Burns (2018). It was therefore assumed the average time was 21.22 minutes.	£77.66 per 21.22 minute consultation (2018)	No	£77.66
GP telephone call	Cost per 1 minute= £3.60	Curtis and Burns (2018)	Average telephone call of 7.1 minutes was taken from Curtis (2015).	£25.99 per 7.1 minute consultation (2018)	No	£25.99
Practice Nurse clinic[1]	£42 per hour with qualifications	Curtis and Burns (2018)	Unit cost used was Practice Nurse with qualifications. This approach was used by Anokye et al. (2018) £42 per hour equates to £0.70 per minute. Time spent with the practice nurse varied from 5-20minutes, therefore a standard published time	£10.85 per 15.5 minute consultation (2018)	No	£10.85

			was applied, assuming the average consultation time is 15.5 minutes, taken from Curtis (2015).			
Physiotherapist	Zero visits baseline nor 12 weeks					
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Chiropractor	Zero visits baseline nor 12 weeks					
Osteopath	Zero visits baseline nor 12 weeks					
Occupational Therapist	Zero visits baseline nor 12 weeks					
Acupuncturist	Zero visits baseline nor 12 weeks					
Specialist Nurse clinic[2]	£87 per hour	Curtis (2017)	Band 7 Specialist Nurse Cost per hour was £87/ cost per minute was £1,45 (Curtis, 2017)	£1.45 (2017)	Yes. UK's GDP Deflator was applied to inflate the price from 2016/17 to 2018/19 (Multiplier 1.0392). £1.51 per minute	£1.51
District Nurse	Zero visits baseline nor 12 weeks					
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Counsellor[1,2]	£49.38 per hour with qualifications	Curtis and Burns (2018)	In line with the other similar allied Health Professionals Band 6 was used (Curtis and Burns, 2018). The cost per working hour is £46 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,410 extra per year (there is no unit cost for Counsellor or Psychologist qualifications, therefore Physiotherapists unit cost	£0.82 per minute, consultation length as reported by participant	No	£0.82

			was used: £5,410/ working hours per year (1,599)= £3.38 per hour. £46+3.38= £49.38 per hour/ £0.82 per minute (£2018).			
Pharmacist [1,2]	£51.17 per hour with qualifications	Curtis and Burns (2018)	It was assumed the Pharmacist would be a Band 6 (Curtis and Burns, 2018). The cost per working hour is £46 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £8,263 extra per year (1,599 hours per year for Band 6): £5.17 per hour £46+5.17= £51.17 per hour/ £0.85per minute (£2018)	£0.85 per minute, consultation length as reported by participant	No	£0.85
Healthcare assistant clinic[2]	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Healthcare Assistants are Band 2 with an annual pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.18 per minute	No	£0.18

Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Support Worker	Zero visits baseline nor 12 weeks					
Phlebotomist[2]	£10.79 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Phlebotomist's are Band 2 with an annual pay rate from April 2018 of £17,260 for those experienced between <1 year to 5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £10.79 per hour or £0.18 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.18 per minute	No	£0.18
Podiatrist	Zero visits baseline nor 12 weeks					
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Dietician[2]	£48.52 per hour with qualifications	Curtis and Burns (2018)	Unit cost is not broken down by type of visit, therefore the same unit cost was applied for clinic visits, home visits and telephone consultations. At Band 6, the average salary Band for Dietician (Curtis and Burns, 2018) the cost per working hour is £45 excluding qualifications (Curtis and Burns,	£0.81 per minute, consultation length as reported by participant	No	£0.81

			2018). With qualifications it is £5,622 extra per year for Dietician. £5,622/ working hours per year (1,599)= £3.52 per hour. £45+£3.52= £48.52 per hour/ £0.81 per minute (£2018).			
Social Worker	Zero visits baseline nor 12 weeks					
Speech Therapist [1,2]	£48.47 per hour with qualifications	Curtis and Burns (2018)	At Band 6, the average salary Band for Speech Therapist (Curtis and Burns, 2018) the cost per working hour is £45 excluding qualifications (Curtis and Burns, 2018). With qualifications it is £5,556 extra per year for Speech Therapist. £5,556/ working hours per year (1,599)= £3.47 per hour. £45+£3.47= £48.47 per hour/ £0.81 per minute (£2018).	£0.81 per minute, consultation length as reported by participant	No	£0.81
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Mental Health Professional[2]	£24.24 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Mental Health Nurses. The participants were not asked to specify what type of Mental Health Professional, therefore it was assumed the same Mental Health Professionals delivering the Behavioural Activation interventions would apply (Curtis and Burns, 2018). Wages of Mental Health	£0.40 per minute	No	£0.40

			Nurses are Band 7 (according to Curtis and Burn, 2018) with an annual pay rate from April 2018 of £38,765 for those experienced between 4-5 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £24.24 per hour or £0.40 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just one participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).			
Health Trainer[2]	£13.65 per hour	Agenda for Change (2018)	According to the Agenda for Change Band's Healthcare Assistants are Band 4 with an annual pay rate from April 2018 of £21,819 for those experienced between 3-4 years. According to Curtis and Burn (2018) most community health care staff work 42.6 weeks (1599 hours) per year. Unit cost= £13.65 per hour or £0.28 per minute. Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).	£0.28 per minute	No	£0.28

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Number of hospital admissions for inpatient care (stayed overnight)	£3,894	NHS reference costs (2018)	Hospital admissions for inpatient care, assuming this care is elective (planned), if it was non-elective inpatient care then this would be a lower cost at £1,603 per case. Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups. Type of procedure in secondary care was only recommended as a bolt on module item to collect for studies specifically concerned with extended hospital care because for instance, admissions and re-admissions are prevalent (Thorn et al. 2018) ISRUM Each HRG had an expected bed day, it was assumed participants did not exceed the expected bed day and so the exceeded bed day unit cost (£346) was not added on.	£3,894	No	£3,894

Cost item	Published unit cost	Published source for unit cost	Assumptions and calculations of unit cost	Calculated unit cost (price year)	Adjustment to price year	Unit cost for £2018
Number of hospital admissions for day patient care	£742	NHS reference costs (2018)	Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups.	£742	No	£742
Number of hospital outpatient appointments	£125	NHS reference costs (2018)	Cost data in the UK is coded by Healthcare Resource Groups (HRGs) for which there are over 2,812 groups (NHS 2018). Selecting a cost requires accurate description by the participant if using a self-reported questionnaire as well as expertise on the descriptions of the HRGs and then judgement to decide which description matches the participants best out of the 2,812 HRG groups.	£125	No	£125
Number of visits to A&E	£160	NHS reference costs (2018)	A&E attendance	£160	No	£160
Number of admissions to hospital, after A&E	£1,603	NHS reference costs (2018)	Non-elective inpatient (excluding excess bed days): £1,603 per case. Non-elective means emergency, but can be via GP not just A&E	£1,603	No	£1,603

Intervention costs						
Cost item	**Published unit cost**	**Published source for unit cost**	**Assumptions and calculations of unit cost**	**Calculated unit cost (price year)**	**Adjustment to price year**	**Unit cost for £2018**
Grade 6 Researcher*	£23.54	LJMU pay (2018)	Grade 6 Researcher (mean annual income of £35,929) including salary on-costs (national insurance and superannuation of 14%). University contract of 6 weeks annual leave, 8 bank holidays, 4 days Christmas closure; 35 hours per week. 1526 hours per year. Hourly unit cost= £23.54	£23.54	No	£23.54
Senior Manager*	£35,000-£50,000	UK average by Search Recruitment study (2010)	Call Centre Manager (mean annual income of £42,500 which excludes salary on costs). If including salary on-costs (national insurance UK 2018 rate of 13.8% and superannuation of 14% was applied in line with the research professionals in this study) the annual total cost=£42,550 x 1.278= £54,315 per year. Assuming senior permanent staff have longer holidays: Assuming a control of 6 weeks annual leave, 8 bank holidays, 4 days Christmas closure; 35 hours per week. 1526 hours per year. Hourly unit cost= £35.59	£35.59	Yes. UK's GDP Deflator was applied to inflate the price from 2010/11 to 2018/19 (Multiplier 1.1424).	£40.66
Team Manager*	£25,000-£30,000	UK average by Search Recruitment study (2010)	Team Manager (mean annual income of £27,500 which excludes salary on costs). If including salary on-costs (national insurance UK 2018 rate of 13.8% and superannuation of 14% was applied in line with the research professionals in this study) the annual total cost=£27,500 x 1.278= £35,145	£23.03	Yes. UK's GDP Deflator was applied to inflate the price from 2010 to 2018/19 (Multiplier 1.1424).	£26.31

			per year. Assuming 28 days (5.6 weeks) leave; 37.5 hours per week. 1740 hours per year. Hourly unit cost= £20.19			
Resource planner*	£18,000-£23,000	UK average by Search Recruitment study (2010)	Call Centre Manager (mean annual income of £20,500 which excludes salary on costs). If including salary on-costs (national insurance UK 2018 rate of 13.8% and superannuation of 14% was applied in line with the research professionals in this study) the annual total cost=£20,500 x 1.278= £26,199 per year. Assuming 28 days (5.6 weeks) leave; 37.5 hours per week. 1740 hours per year. Hourly unit cost= £15.05	£15.05	Yes. UK's GDP Deflator was applied to inflate the price from 2010 to 2018/19 (Multiplier 1.1424).	£17.19
Call agent*	£12,000-£16,000	UK average by Search Recruitment study (2010)	Call Centre Manager (mean annual income of £12,000 which excludes salary on costs, over 80% agents are not permanent staff; majority of staff had worked there less than 1 year). Assuming 28 days (5.6 weeks) leave; 37.5 hours per week. 1740 hours per year. Hourly unit cost= £6.89	£6.89	Yes. UK's GDP Deflator was applied to inflate the price from 2010 to 2018/19 (Multiplier 1.1424).	£7.87
Travel to contact centre	0.15 pence per mile	AA mileage calculator (2018)	7.4 miles one way (14.8 miles return) based on postcode of research institute and contact centre postcode. One car for all researchers.	£2.22	No	£2.22
Posturite height adjustable desk	£319.14	Posturite deskrite 100 sit-stand platform (2018)	Inclusive of VAT. https://www.posturite.co.uk/deskrite-100-sit-stand-platform.html The desk life was not reported. It was assumed to have an expected lifetime of 5 years (as recommended in Drummond et al. (2015) on the expected lifetime of equipment). Calculation: 52	£14.73	No	£14.73

			weeks x 5= 260 weeks. Cost for 12 weeks= 12/260= 0.046 x £319.14= £14.73 per 12 week period			
Laminated instructions card	£0.07 per A4 page in colour; Lamination £1.04 per A4 page	Research Insitute's Printing costs (2018)	Calculation per participant: £0.07+£1.04= £1.11	£1.11	No	£1.11
Timer	£2.38	Receipt (2018)	Unit cost per timer: £2.38 inclusive of VAT https://www.nisbets.co.uk/canteen-magnetic-countdown-timer/df672	£2.38	No	£2.38
Log book	£0.07 per page in colour	Research Insitute's Printing costs (2018)	Calculation per participant: A7 booklets requires 2 pages, £0.07 x 2= £0.14	£0.14	No	£0.14

* Efforts to estimate the overhead costs were not done due to the small magnitude of this cost in terms of the quantity and unit cost (just two participant reported this resource) and the small magnitude of the unit cost for this health professional (Drummond et al. 2015: 220).

Appendix D.4. Disaggregated costs and consequences

Table 1. Cost-utility analysis results unadjusted results

Variable	SLaMM+	SLaMM
	Mean (SD) per participant	Mean (SD) per participant
QALYs	0.188 (SD: 0.045)	0.201 (SD: 0.031)
Total costs	£497.51 (SD: £1,356.59)	£309.22 (SD: £89.16)
Incremental QALYs & Costs		
Incremental QALYs: SLaMM+ vs SLaMM	-0.0137 (SE: 0.0124; 95% CI: -0.0381 to 0.0106); this means the SLaMM+ group has 0.0137 less QALYs compared to SLaMM	
Incremental Costs: SLaMM+ vs SLaMM	£188.29 (SE:£318.81; 95% CI: £-436.57 to £813.15); this means the SLaMM+ group costs £188.29 more than the SLaMM group	
ICER statistic		
ICER	£-13,731 per QALY	

Table 2. Research Institute (Payer) Perspective: Intervention operating costs

Key intervention activities	SLaMM+		SLaMM	
	Description of resources consumed	Mean cost per participant	Description of resources consumed	Mean cost per participant
Height-adjustable desk	Cost of height-adjustable desk for 12-weeks assuming a desk life of 5 years; Printing; Laminated instruction sheet	£14.23	n/a	£0
Weekly emails	Time of Grade 6 researcher to disseminate infographic via email (5mins x12 weeks)	£0.59	Time of Grade 6 researcher to disseminate infographic via email (5mins x12 weeks)	£0.59
Organisation of Education & Training Sessions	Time of Grade 6 researcher to organise sessions (15 mins x3 sessions)	£0.44	Time of Grade 6 researcher to organise sessions (15 mins x3 sessions)	£0.44
Delivery of Education & Training sessions	Time of Grade 6 researcher (30mins x 3+15mins preparation time); Travel costs to workplace (14.8 miles return journey); Timer; Printing	£7.80	Time of Grade 6 researcher (30mins x 3+15mins preparation time); Travel costs to workplace (14.8 miles return journey); Timer; Printing	£7.80
Total costs		**£23.06**		**£8.83**

Table 3. Employer (host) Perspective: Intervention operating costs

Key intervention activities	SLaMM+		SLaMM	
	Description of resources consumed	Mean cost per participant	Description of resources consumed	Mean cost per participant
Height-adjustable desk	No costs incurred by employer	£0	n/a	£0
Weekly emails	Time of Centre Contact to disseminate via email (5mins x12 weeks); Time of call agent to review infographic via work email (1minx12 weeks)	£2.23	Time of Centre Contact to disseminate via email (5mins x12 weeks); Time of call agent to review infographic via work email (1minx12 weeks)	£2.23
Organisation of Education & Training Sessions	Time of Centre Contact to organise sessions (15 mins x3 sessions); Time of Resource Planner to organise sessions (3x45minutes to scheduling offline time+15minutes email/phone communication)	£2.92	Time of Centre Contact to organise sessions (15 mins x3 sessions); Time of Resource Planner to organise sessions (3x45minutes to scheduling offline time+15minutes email/phone communication)	£2.92
Delivery of Education & Training sessions	Time of agents to attend sessions during worktime (30minutesx3 sessions); Time of Centre Contact to coordinate session (3x15minutes).	£12.30	Time of agents to attend sessions during worktime (30minutesx3 sessions); Time of Centre Contact to coordinate session (3x15minutes).	£12.30
Total costs		**£17.45**		**£17.45**

Table 4. Research Institute Perspective: Intervention set up costs

Key intervention activities	SLaMM+		SLaMM	
	Description of resources consumed	Mean cost per participant	Description of resources consumed	Mean cost per participant
1 hour consultation with senior management (approximately 3 months before intervention starts)	Time of two Grade 6 researchers (1 hour); Printing; Travel costs to workplace (14.8 miles return journey).	£1.27	Time of two Grade 6 researchers (1 hour); Printing; Travel costs to workplace (14.8 miles return journey).	£1.27
30 minute consultation with Resource Planner (approximately 3	Time of two Grade 6 researchers (30 minutes); Printing; Travel costs to workplace (14.8 miles return journey).	£0.66	Time of two Grade 6 researchers (30 minutes); Printing; Travel costs to workplace (14.8 miles return journey).	£0.66

months before intervention starts)				
Organisation of Team Manager briefs (approximately 3 months before intervention starts)	Time of Grade 6 researchers via phone/ email (15 minutes).	£0.15	Time of Grade 6 researchers via phone/ email (15 minutes).	£0.15
Delivery of Team Manager briefs (approximately 2 months before intervention starts)	Time of Grade 6 Researcher (45 minutes including preparation time); Travel costs to workplace (14.8 miles return journey).	£0.50	Time of Grade 6 Researcher (45 minutes including preparation time); Travel costs to workplace (14.8 miles return journey).	£0.50
Organisation of Agents briefs (approximately 2 months before intervention starts)	Time of Grade 6 researchers via phone/ email (15 minutes)	£0.15	Time of Grade 6 researchers via phone/ email (15 minutes)	£0.15
Delivery of Agent briefs (approximately 1 month before intervention starts)	Time of Grade 6 researchers to deliver briefs (15 minutes per session)	£0.64	Time of Grade 6 researchers to deliver briefs (15 minutes per session)	£0.64
Installation of height-adjustable desk (1 working day before intervention start date)	Time of two Grade 6 researcher to install height-adjustable desks (2 hours); Travel costs to workplace (14.8 miles return journey).	£2.41	n/a	£0
Total cost per participant		**£5.77**		**£3.36**

Table 5. Employer's Perspective: Intervention set up costs

Key intervention activities	SLaMM+		SLaMM	
	Description of resources consumed	Mean cost per participant	Description of resources consumed	Mean cost per participant
1 hour consultation with senior management (approximately 3 months before intervention starts)	Time of Senior Manager at company (1 hour); Time of Centre Contact (1 hour)	£1.67	Time of Senior Manager at company (1 hour); Time of Centre Contact (1 hour)	£1.67
30 minute consultation with Resource Planner (approximately 3	Time of Senior Manager at company (30 minutes); Time of Resource Planner (30 minutes)	£1.26	Time of Senior Manager at company (30 minutes); Time of Resource Planner (30 minutes)	£1.26

months before intervention starts)				
Organisation of Team Manager briefs (approximately 3 months before intervention starts)	Time of Resource Planner to organise sessions (45 minutes); Time of Centre Contact to organise sessions via phone/ email (15 minutes)	£0.53	Time of Resource Planner to organise sessions (45 minutes); Time of Centre Contact to organise sessions via phone/ email (15 minutes)	£0.53
Delivery of Team Manager briefs (approximately 2 months before intervention starts)	Time of two Centre Contacts to coordinate sessions (15 minutes); Time of 13 Team Managers to attend (15 minutes)	£2.47	Time of two Centre Contacts to coordinate sessions (15 minutes); Time of 13 Team Managers to attend (15 minutes)	£2.47
Organisation of Agents briefs (approximately 2 months before intervention starts)	Time of Resource Planner to organise sessions (45 minutes); Time of Centre Contact to organise sessions via phone/ email (15 minutes)	£1.07	Time of Resource Planner to organise sessions (45 minutes); Time of Centre Contact to organise sessions via phone/ email (15 minutes)	£1.07
Delivery of Agent briefs (approximately 1 month before intervention starts)	Time of agents to attend brief (15 minutes); Time of two Centre Contacts to coordinate sessions (15 minutes).	£2.26	Time of agents to attend brief (15 minutes); Time of two Centre Contacts to coordinate sessions (15 minutes).	£2.26
Installation of height-adjustable desk (1 working day before intervention start date)	Time of two Centre Contacts to coordinate the installation of the height-adjustable desks (5 hours)	£6.51		£0
Total costs		**£15.77**		**£9.26**

Table 6. Most commonly used primary care activity

Resource type	Unit Cost	SLaMM+		SLaMM	
		Mean number of visits (SD); Number of people reporting any use	Total Mean Cost (SD)	Mean number of visits (SD); Number of people reporting any use	Total Mean Cost (SD)
GP consultations at Baseline	£34 per 9.11 minute consultation	1.17 (2.20); n=9/19	£38.78 (£73.07)	1.95 (2.42); n=13/21	£71.09 (£93.26)
GP consultations at 6 months		1.67 (2.74); n=8/19	£54.44 (£88.08)	1.76 (3.87); n=12/21	£59.10 (£113.72)

Practice Nurse consultations at Baseline	£10.85 per 15.5 minute consultation	0.15 (0.50); n=2/19	£1.71 (£5.44)	0.28 (0.78); n=3/21	£3.10 (£8.50)
Practice Nurse consultations at 6 months		0.21 (0.71); n=2/19	£2.28 (£7.74)	0.19 (0.51); n=3/21	£2.07 (£5.55)
Counsellor at Baseline	£0.82 per minute visit, range of average visit was 45 minutes	1.05 (3.34); n=0/19	£0 (£0)	1.05 (3.34); n=0/21	£0 (£0)
Counsellor at 6 months		1.05 (3.34); n=3/19	£38.84 (£123.29)	0 (0); n=0/21	£0 (£0)

Table 7. Breakdown of all secondary care activity

Resource type	Unit Cost	SLaMM+		SLaMM	
		Number of people reporting any use	Total Mean Cost (SD)	Number of people reporting any use	Total Mean Cost (SD)
Outpatient visit at Baseline	£125 per visit	n=2/19	£19.74 (£62.68)	n=4/21	£41.67 (£99.48)
Outpatient and Day Case visit at 6 months		n=3/19	£39.47 (£102.51)	n=3/21	£35.71 (£89.64)
Day case visit at Baseline	£742 per visit	n=0/19	£0	n=1/21	£35.33 (£161.92)
Day case visit at 6 months		n=1/19	£78.11 (£340.45)	n=0/21	£0
Inpatient admission at Baseline	£3,894 per admission	n=0/19	£0	n=1/21	£185.43 (£849.74)
Inpatient admission at 6 months		n=1/19	£204.95 (£893.34)	n=1/21	£185.43 (£849.74)
A&E visits at Baseline	£160 per visit	n=2/19	£16.84 (£50.45)	n=1/21	£7.62 (£34.91)
A&E visits at 6 months		n=1/19	£33.68 (£146.83)	n=1/21	£7.62 (£34.91)

Table 8. Prescribed medications in the last 12 weeks

	SLaMM+	SLaMM	
Prescribed any medication	Baseline 40% (n=6/15); 12-weeks 60% (n=9/15)	Baseline 38.89% (n=7/18); 12-weeks 44.44% (n=8/18)	+14.45%
Prescribed moderate-to-strong painkillers	Baseline 13.33% (n=2/15); 12-weeks 13.33% (n=2/15)	Baseline 22.22% (n=4/18); 16.66% 12-weeks (n=3/18);	-5.56%
Prescribed antibiotics	Baseline 6.67% (n=1/15); 12-weeks 0% (n=0/15)	Baseline 11.11% (n=2/18); 12-weeks 11.11% (n=2/18)	+6.67%
High blood pressure	Baseline n=1/15; 12 weeks n=1/15	Baseline n=0/18; 12 weeks n=0/18	No change
High cholesterol	Baseline n=1/15; 12 weeks n=1/15	Baseline n=0/18; 12 weeks n=0/18	No change
T2D	Baseline n=1/15; 12 weeks n=1/15	Baseline n=0/18; 12 weeks n=0/18	No change
Stomach	Baseline n=1/15; 12 weeks n=1/15	Baseline n=0/18; 12 weeks n=0/18	No change
Weight loss for obesity	Baseline n=1/15; 12 weeks n=1/15	Baseline n=0/18; 12 weeks n=0/18	No change
Antidepressants	Baseline n=1/15; 12 weeks n=1/15	Baseline 5.56% n=1/18; 12 weeks 0% n=0/18	-5.56%
Anti-inflammatory/ allergy	Baseline 13.33% n=2/15; 6.67% 12 weeks n=1/15	Baseline 0% n=0/18; 12 weeks 5.56% n=1/18	+12.23%

Table 9. Productivity loss

Employer's costs			
Variable	SLaMM+ (n=16/19)	SLaMM (n=14/21)	Difference
Days of certified sickness (Baseline)	21 (n=1/16)	0 (n=0/14)	+21
Days of certified sickness (12 weeks)	6 (n=1/16)	11 (n=1/14)	-5
Employee's costs			
Variable	SLaMM+ (n=16/19)	SLaMM (n=14/21)	Mean difference
Hours of uncertified sickness (Baseline)	Mean: 15hr32min34secs Median: 15hrs20mins00secs (n=14/16)	Mean: 16hr47min13secs Median: 5hrs31mins30secs (n=9/14)	Mean: -1hr14mins39secs (less for SLaMM+) Median: +9hr48mins30secs (more for SLaMM+)
Hours of uncertified sickness (12 weeks)	Mean: 11hr37min19secs	Mean: 7hr42min43secs	Mean: +3hr54mins36secs (more for SLaMM+)

	Median: 6hrs30mins00secs (n=11/16)	Median: 7hrs03mins00secs (n=10/14)	Median: +33mins30secs (less for SLaMM+)
Employer or employee cost*			
Agents with 4 or more days of uncertified sickness who may have claimed Statutory Sick Pay (Baseline)	n=1/16	n=0/14	+1
Agents with 4 or more days of uncertified sickness who may have claimed Statutory Sick Pay working days (12 weeks)	n=3/16	n=0/14	+3
Agents with 2-3 days uncertified sickness who may have claimed Statutory Sick Pay for working and non-working days (Baseline)**	n=2/16	n=3/14	-1
Agents with 2-3 days uncertified sickness who may have claimed Statutory Sick Pay for working and non-working days (12 weeks)**	n=1/16	n=2/14	-1

*This depended upon whether the agent had been working at the company for a minimum of 3 months. If they had worked there for less than 3 months they had to claim from the state; **this depended on whether the employee claimed sickness during the weekend days (non-working days)

Table 7. QALY breakdown by group (unadjusted)

	SLaMM+	SLaMM
	Mean (SD)	Mean (SD)
Life years	0.23 years/ 12 weeks	0.23 years/ 12 weeks
Unadjusted HRQoL for all three time points	0.814 (0.194)	0.873 (0.135)
HRQoL at Baseline	0.780 (0.212)	0.884 (0.154)
HRQoL at 12 weeks	0.847 (0.220)	0.862 (0.137)

Appendix E.1. Microcosting tool

1. People (Human resources)	Type of resource and context: 1. What is your job title and if possible skill level/ grade? 2. Name the tasks (activities) which you do which are related to the intervention? *(if they perform more than one task, perform this questioning exercise for each task starting in chronological order)* 3. When do you perform this task? *(before the intervention starts, first week, after the intervention)* 4. Who else is involved in this task?* Quantification of resource: 5. On average how long does this task take to perform each time? *(duration)* 6. How many times is this take performed? *(frequency)* Additional context to consider opportunity costt: 7. Where any specific tasks (activities) given up or done differently because of this additional task? 8. Was the impact minor, moderate or major? *This can include the participant. Arrange a consultation with the people identified in question 4 and repeat questions 1-8 with those people. If it is anticipated that it will not be feasible to arrange a consultation with any of the other people identified, ask the present stakeholder to estimate their resource use by answering question 1-7 on their behalf.
2. Place/ Setting (Capital resources)	Quantification of capital resources described in natural units: 1. How is this task performed? *(e.g. face-to-face, email, telephone)* 2. Where does the task take place? *(private room at leisure centre)* 3. How many times is this task performed in this place? *(frequency)* 4. On average how much time is required to perform the task in this place? *(duration)* 5. If you have to travel to this place, how much of your time does it take to travel one way? 6. On average what mode of transport do you use? 7. Which other places, if any, does this task take place at?* Additional context to consider opportunity cost: 8. Where any specific activities not related to the intervention given up or done differently because of the place being used for this task?

	9. Was the impact minor, moderate or major? *Repeat questions 1-7 for each place identified for the task
3. Materials (Equipment)	Quantification of equipment described in natural units: 1. What materials are required for this task? 2. On average how much of this specific material is required for this task? *(frequency)* 3. Which other equipment, if any, does this task take place at?* Additional context to consider opportunity costt: 4. Where any specific activities not related to the intervention given up or done differently because of the equipment being used for this task? 9. Was the impact minor, moderate or major? *Repeat questions 1-5 for each equipment identified for the task
4. Out-of-pocket payments	Quantification of out-of-pocket expenses: 1. What out-of-pocket payments are paid for this task (exclude payments for any capital resources or equipment)? 2. How many times is this paid? 3. What other out-of-pocket payments, if any are paid?* Additional context to consider opportunity cost: 4. Is anything, not related to the intervention, given up or done differently because of this out-of-pocket payment? 5. Was the impact minor, moderate or major? *Repeat questions 1-5 for each out-of-pocket payment identified for the task

Appendix E.2. Example of resource use quantification

Task (activity): Scheduling of Education & Training sessions		
Context	Calculations to quantify resource use	**Total resource use (per intervention participant)**
1. *Human Resource:* Resource Planner (junior level) schedules three Education and Training sessions on weeks 1, 3 and 7 of the SLaMM intervention for 60 participants. They also liase with the researchers via telephone and email. Minor impact, no specific activities given up or done differently.	• 3 x 45 minutes of scheduling per 60 participants • 3 x 15 minutes of liaising with researchers per 60 participants	**3 minutes**
2. *Place:* In usual office, on usual computer and via telephone to speak to researchers. The schedule for education and training session was via email to the researcher. No other place involved. Minor impact, no specific activities given up or done differently (negligible resource use).	• 3 x IT system • 3 x Telephone system	**0**
3. *Materials:* None	Not applicable	**0**
4. *Out of pocket payments:* None	Not applicable	**0**

Appendix E.3. Example of presentation of results

Stakeholder (perspective): Leisure Centre					
Activity	Resource type	Resource item	Intervention A	Intervention B	**Key Result: Incremental Analysis (per participant)**
			Total resource use (per participant)*	Total resource use (per participant)*	
Education session	Human Resource	Time of Gym Instructor (one to one)	30 minutes	15 minutes	**+15 minutes**
	Capital	Consultation room	30 minutes	15 minutes	**+ 15 minutes**
	Material	Information leaflet in colour	6 pages	2 pages	**+ 4 pages**
Gym class	Human Resource	Time of Gym Instructor	3 minutes	Not applicable	**+3 minutes**
	Capital	Gym class room (15 people: 45 min class)	3 minutes	Not applicable	**+3 minutes**
Stakeholder (perspective): Healthcare Sector					
Activity	Resource type	Resource item	Intervention A	Intervention B	**Key Result: Incremental Analysis (per participant)**
			Total resource use (per participant)*	Total resource use (per participant)*	
Primary care	Human Resource	GP visits	5 visits	2 visits	**+3 visits**
		Practice Nurse visits	2 visits	2 visits	**No difference**
Secondary care	Services	A&E visits	1 visits	0 visits	**+1 visits**
		Outpatient appointments	1 visits	4 visits	**-3 visits**

*Report means