Economic evaluation of universal programmes to promote healthy behaviours: challenges and possible solutions with an application to physical activity

Francesco Paolo Candio

University of Leeds

School of Medicine

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

June 2019

Declaration

I confirm that the work submitted is my own and that appropriate credit has been
given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and
that no quotation from the thesis may be published without proper acknowledgement.

Francesco Paolo Candio

June 2019

Acknowledgements

Abstract

Several methodological challenges characterise the economic evaluation of public health interventions, especially universal programmes to promote healthy behaviours. While relevant guidance to support analysts exists, methodological shortcomings have been highlighted in recent reviews. The choice of evaluation method and assumptions can have profound impacts on cost-effectiveness, and consequently decision-making. Therefore, methods development and validation for public health interventions are needed.

Narrowing the focus to physical activity, the first research question of this thesis was: if and how have the methodological challenges acknowledged in public health have been addressed in economic evaluations of PA promotion interventions? To address this question, a systematic review of the methods used for economic evaluation was conducted. This review revealed a paucity in the methods used and identified four outstanding issues to be addressed in this thesis: modelling of heterogeneity and inequality in population-level impact, assumptions regarding maintenance of behaviour change over time, longitudinal selection bias and perspective for economic evaluation. To address these four issues, this thesis was framed around two research questions: how to incorporate concerns regarding population-level impacts into the economic model? What is the impact of the key methodological assumptions that underpin the existing models on the economic decision? These questions are explored using a case study which focused on a local universal programme to promote physical activity.

A novel modelling approach is proposed to address the issues of population-level impact and behaviour change maintenance over time. Results show the choice of evaluation method can impact the cost-effectiveness decision. In particular, the choice of method and assumptions regarding selection mechanisms and maintenance of behaviour change over time have the potential to independently affect identification of the optimal strategy. Results also suggest that the case study programme can potentially be cost-neutral, but widen existing health inequalities in the short term, if its opportunity cost is considered.

Findings highlight the importance of addressing the identified shortcomings to adequately inform decision-making. They provide support for change in implementation and research practices and pave the way to more robust and informative economic evaluations. The proposed modelling approach represents a simple modelling solution that can be replicated for evaluation of universal strategies to promote healthy behaviours. A summary of the lessons learnt, which could be useful for future evaluations, is presented.

Keywords: economic evaluation, healthy behaviours, physical activity, public health economic modelling, universal promotion.

Table of Contents

List of Figures

List of Tables

Table of Acronyms and Abbreviations

	Expansion
BMI	Body Mass Index
CBA	Cost-Benefit Analysis
CCA	Cost-Consequences Analysis
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CUA	Cost-Utility Analysis
DALY	Disability-Adjusted Life Year
DCEA	Distributional Cost-Effectiveness Analysis
EE	Economic Evaluation
EMC	Embedded Markov Chain
HWB	Health and Well-being Board
ICER	Incremental Cost-Effectiveness Ratio
IMD	Index of Multiple Deprivation
INMB	Incremental Net Monetary Benefit
LLGA	Leeds Let's Get Active
LY	Life-Year
MET	Metabolic Equivalent of Task
MAR	Missing At Random
MC	Markov Chain
MCAR	Missing Completely At Random
MNAR	Missing Not At Random
NAD	Number of Active Days
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PA	Physical Activity
RCT	Randomised Controlled Trial
QALE	Quality-Adjusted Life Expectancy
QALY	Quality-Adjusted Life Year
RR	Relative Risk

1 Chapter

Introduction

1.1 Background

1.1.1 What is economic evaluation?

Economic evaluation (EE) is a type of economic analysis which compares at least two alternative and mutually exclusive investment options in terms of their costs and consequences (Drummond, 2015). As with all other applications in economics, EE is needed because there is scarcity of resources, relative to wants or needs. With scarcity, rationing becomes unavoidable.

Due to market failures that characterise the health care sector, government involvement and non-market methods are required to ensure efficient resource allocation. These failures include: 1) the uncertainty surrounding the need for health care (i.e. we do not know when we are going to need it); 2) asymmetric information between providers and consumers of health care (i.e. moral hazard); 3) monopoly / barriers to provision: especially for acute care services in vulnerable groups, the proximal hospital becomes the only provider, which is regulated by professional licensure; 4) externalities: there represent spill-over effects from other consumers of health care (e.g. vaccination, which benefit the community as a whole due to herd immunity); 5) equity: society values fair treatments for all citizens. (Drummond, 2015). In order to support decision-makers in health care, EE has been used

1

increasingly to address optimisation problems under budget constraints (Briggs, 2006).

In order to inform these decisions, incremental analysis provides a way to account for the value of the next best alternative forgone (i.e. opportunity cost) and identify the most economically efficient option (Drummond, 2015). To this purpose, partial and full EE frameworks are available. With partial EE, the comparison is limited to the costs of the alternatives being assessed. Full EE instead requires comparing the options both in terms of costs and consequences.

Differences between principles for guiding decision-making in health care have been debated for decades. Contrasts between positions have arisen not simply due to technical issues, many of which are confronted with any form of EE (e.g. in obtaining precise estimates of costs and effect differences), but rather fundamental questions regarding the theoretical foundations of different approaches to social welfare (Gray et al., 2011).

In economics, the concept of utility has been traditionally used as a measure of pleasure or satisfaction that consumers obtain from goods and services. Within society, the level of utility (welfare) derived, from a particular choice, varies from individual to individual, according to individual preferences (Bentham, 2000).

The two most debated theoretical paradigms regarding social welfare are known as "welfarism" and "extra-welfarism". As argued by Brouwer et al. (2008), at the root of the distinction between these two paradigms there are two interrelated topics: the source and nature of valuation and the Pareto principle. In welfarism, social welfare is seen as the sum of individual utilities which, in turn, can be used to measure consumers' pleasure and satisfaction obtained from any good and service (including health). Under this school of thought, the initial distribution of income and wealth is taken as a given and changes in either ought to satisfy the Pareto principle. That is, social welfare increases if the welfare of any member of society increases and the welfare of all others remains the same. Or if the welfare gains from one individual outweigh the losses incurred by other individuals (i.e. potential Pareto principle). In other words, losers are compensated so that no one is worse off after implementation of the allocative decision.

Extra-welfarism rejects a welfarism approach which is based merely on individual utility, to broaden the evaluative space to other criteria. The individual utility information is not discarded, rather it is complemented by other information including, for example, equity to reach a "quasi-utilitarian" balance Brouwer et al. (2008). From such perspective, uncompensated changes that would not satisfy the Pareto criterion may be judged to be social deteriorations (or improvements) by invoking additional ethical criteria (e.g. distribution of need).

Both paradigms have weaknesses. Extra-welfarism imposes methodological limitations, in particular, that of not considering costs and outcomes impacting other sectors ("goods") which may relevant to society. Furthermore, a main underlying assumption of this stance is that everyone values health states similarly.

More recently extra-welfarism has prevailed in terms of applied work within the health care context, with methodological frameworks derived from this paradigm outnumbering those from welfarism (Weatherly et al., 2009). This has been especially true within public health settings, due to the capacity of an extra-welfarism framework to incorporate concerns regarding health equity which is of primary importance to these decision-makers. One of the most remarkable examples of such type of frameworks is that of Amartya Sen's capabilities (Robeyns, 2016). This theoretical framework entails that freedom to achieve well-being is a fundamental tenet to society and needs to be understood in terms of people's real opportunities to achieve functionings. In other words, freedom or valuable opportunities that anyone can choose from to do and be what they have reason to value. Culyer in part applied Sen's capabilities approach to develop an extra-welfarist perspective whereby health status directly influences social state preferences (Culyer, 1990). Culyer's theoretical work provided the basis for using quality-adjusted life-years for evaluation by the UK National Institute of Health and Clinical Excellence (NICE, Coast et al. 2008).

1.1.1.1   Types of economic evaluation

Four basic forms of full EE can be employed, which differ in the nature and way consequences are considered (Drummond, 2015).

4

Cost-consequences analysis

In cost-consequences analysis (CCA), costs and outcomes are presented in a disaggregated tabular or graphical format. As a result, decision-makers are left to form their own opinion on relevance and importance of the outputs. CCAs have been recommended for interventions that generate an array of health and non-health effects which are difficult to combine in a single metric (National Institute for Health and Care Excellence, 2012). In fact, CCA assumes that the decision-maker are able to reliably and consistently process such information and should put their own weight on the different outcomes.

Cost-effectiveness analysis

Rooted in an extra-welfarism approach, in cost-effectiveness analysis (CEA), outcomes are measured in natural effects (e.g. life-years, LYs) or physical units and results are presented in aggregated formats. Consequences are not valued formally, but their identification implies at least relevance, if not importance of achieving them. CEA allows for comparison across interventions that focus on the same outcome, with results being expressed in the form of cost-effectiveness ratios (e.g. £ 5,000 per disease case averted). However, CEA lacks ability to compare different interventions across the health care sector.

Cost-utility analysis

Based on an extra-welfarism perspective, cost-utility analysis (CUA) can be considered as a particular form of CEA, where the considered outcome is (health) utility. In health care and public health settings, the QALY represents the most commonly used preference-based metric (Weatherly et al., 2009), which captures both survival and quality of life measures. Health outcomes are measured in terms of quality of life, using a scale from 0 to 1 (0=death and 1=full health), to be combined with duration of life (survival, Gray et al., 2011).

In order to generate a value for quality of life, particular health states are valued according to preferences (unlike natural units). In practical terms, such valuation can be conducted using subjective judgements, direct elicitation methods (e.g. visual analogue scales, time trade-offs and standard gamble methods (Craig et al. 2009) or using multi-attribute utility scales (e.g. SF-6D, EQ-5D (Devlin and Brooks, 2017). In the case of EQ-5D, health state valuations were compiled from responses of a large sample of the British adult population (Dolan, 1997). Tariff weights that can be applied to each level of the attributes were thus developed for conversion to an overall single value (from 0 to 1 of utility) for each EQ-5D state.

One of the key advantages of using a generic measure of health is that it allows for comparison across all the entire health sector (Drummond, 2015). However, among other aspects, the QALY has been critised for its limited ability to capture relevant health effects, in particular mental health and wellbeing (Philips, 2009). This was subsequently acknowledged by the

Department of Health who commissioned research on subjective measures of wellbeing (Bache and Reardon, 2016).

Cost-benefit analysis

Based on a welfarist perspective, cost-benefit analysis (CBA) is potentially the broadest form of analysis that can potentially overcome some of the limitations of the other analytical frameworks. Consequences are valued in monetary terms, therefore, evaluation of the value for money of an alternative becomes straightforward. CBA also allows for comparisons not just within the health care budget, but also different areas of the public sector (e.g. environment, criminal justice, education).

However, this form of EE has seen limited application in health care (Weatherly et al., 2009). This can be largely attributed to the difficulty of measuring (McIntosh, E. and Clarke, P. 2010) and aversion of placing monetary values to health states (Cookson et al. 2008). Two main methods have been used for the former: the human capital and the willingness to pay approaches. In the human capital approach, the present value of the individual's future earnings is used as the measure of the value of losses and gains from premature mortality or morbidity states. The latter approach instead uses revealed preferences (from past behaviour) or stated preferences (surveys) or contingent valuation exercises (McIntosh, E. and Clarke, P. 2010).

1.1.2 What is health promotion?

Health promotion is a branch of public health. Acheson (1988) defined public health as "*the art and science of preventing disease, prolonging life and promoting health through the organized efforts of society*". While public health overlaps to some extent with health care in practice (e.g. disease management), health care focuses on the treatment of ill health of the *individual*, while public health on the determinants of health through preventive actions directed at *population* level. Health promotion has been defined in its First International Conference in Ottawa (World Health Organization, 1986) as the process of enabling people to increase control over their own health, by gaining control over the underlying determinants of health (World Health Organization, 2016).

Health promotion activities

Throughout this thesis, the words "intervention", "initiative", "programme", "activity" and "policy" are used interchangeably to define any organised action aimed to modify the determinants of health and health inequality. Especially when directed at entire populations, health promotion activities typically aim to modify the conditions in which people live or behave, by manipulating environmental-level determinants of health and/or health inequality. Throughout this thesis, the words "intervention", "initiative", "programme" and "policy" are used interchangeably to define any organised action aimed to modify the determinants of health and health inequality. These programmes act on longer causal chains to health, compared to clinical interventions. Figure 1.1 provides a simplified illustration of the logical

pathway of these programmes to health improvement (e.g. life expectancy and quality of life).

Figure 1.1 Logical pathway to health improvement
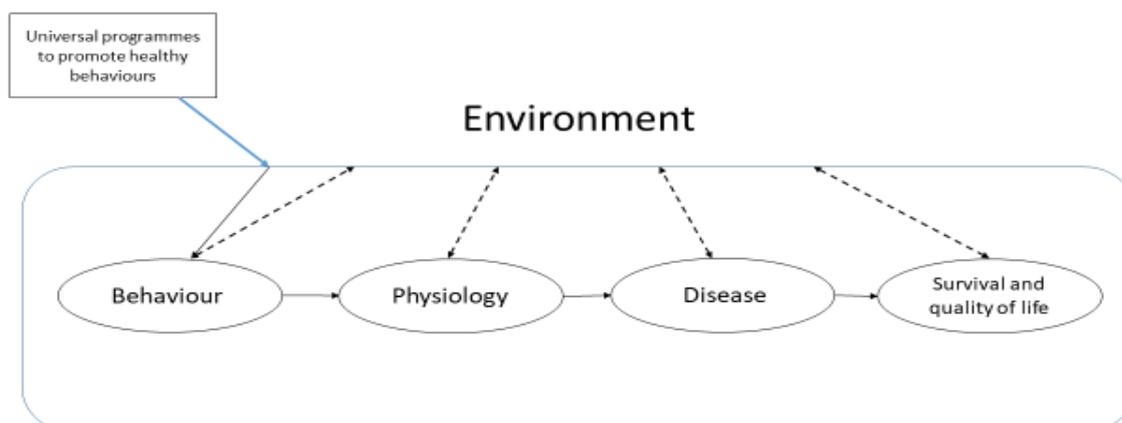
Health promotion actions can differ by country and organisations. Since April 2013 in the UK, following the introduction of the Health and Care Act (UK Government, 2012), responsibility for the promotion of health in the population has been transferred from the National Health Service (NHS) to Local Authorities. To address the rise of non-communicable disease, which have been attributed to lifestyle factors to a large degree (World Health Organization, 2013), the promotion of healthy behaviours has become a priority on the agenda of public health agencies.

Economics of health promotion

The economic case to invest in health promotion is stronger than ever (World Health Organization, 2015). Investment in health promotion can help: improve population health by reducing the incidence and severity of chronic diseases, reduce health inequalities and public spending pressure by

reducing demand for health care and services (NHS Scotland, 2016). Despite this, only a limited proportion of the public spending is allocated to health promotion (The King's Fund, 2018).

Normative reasons aside, the robustness of the methods used to assess the value for money of health promotion initiatives have been challenged (Weatherly et al., 2009), hence the need for further methodological research. In particular, four main macro areas, which are interrelated with each other, have been identified: attribution of effects, measuring and valuing outcomes, inter-sectoral costs and consequences and equity considerations.

In part the issues arise from practical constraints in designing health promotion initiatives (see section 1.1.3). However, particular emphasis has been put on a lack of robust statistical approaches to measure intervention effects and narrow economic evaluation perspectives which limited the ability of studies to capture economic effects relevant to public health decision-making. Specifically, only in a minority of cases have economic evaluations been found to apply appropriate methods to deal with non-experimental data such as matching techniques and to account for relevant costs and outcomes occurring beyond the end of the trial or falling outside of the health care sector (e.g. impact of substance abuse prevention on criminal justice). In addition, none of the economic models were designed to take into account distributional health effects, despite health equity being the primary aim of the intervention and a key public health objective. A few attempted to explore this informally, for example through discussion of implications.

1.1.2.1 Theoretical foundations

Within the broader sphere of public health, health promotion agendas have been characterised by theoretical discussions on the topics of health equity and behaviour change. In this literature, debate has taken place on the reasons why health inequalities exists, why especially individuals from deprived socio-economic backgrounds may find it difficult to live healthy lifestyles and, consequently, how to promote healthy behaviours. While a thorough discussion of these matters would go beyond the scope of this thesis, an overview of the main paradigms and theories in these two fields are presented below. This is to provide theoretical justifications as to why to incorporate concerns related to health equity and behaviour change into the EE of intervention to promote healthy behaviours.

1.1.2.1.1 Health equity

Social disparities in health is a key driving factor for public health and health promotion. Among the various forms of inequalities, the phenomenon known as the social gradient has received most attention (Hart, 1971, Marmot, 2001, Marmot, 2010, Wanless, 2003).

The social gradient refers to a graded relation between socio-economic position and health outcomes, which runs across the whole socio-economic spectrum (Marmot, 2005). In other words, the higher the social position the better the heath. Since the Whitehall II study (Marmot et al., 1991) which found a strong negative association between grade level of employment and mortality rate among 18,000 British civil servants, several studies have

investigated this phenomenon and found consistent results (Asaria et al., 2016a, Marmot, 2010).

What has been less uncontroversial, however, is the root causes of the social gradient and related judgement of fairness of health inequalities (Townsend P., 1988). The Black Report, and a range of subsequent studies, have documented the magnitude and trend of this phenomenon in the UK and proposed alternative explanations for its occurrence Marmot (2010).

Townsend P. (1988) put forward four main types of theoretical explanations of the relationship between health and inequality: artefact, selection, cultural / behavioral and material. The normative debate, which has mostly been centered on the latter two explanations, has taken place primarily within contexts of countries that have passed through what is referred to as the "epidemiological transition". This transition represents the shift of a country from epidemic diseases of poverty, such as infectious diseases, to chronic conditions being the major causes of mortality (Wilkinson, 1994).

While being presented as mutually exclusive explanations, materialist and psychosocial hypotheses often operate together in different combinations and have been integrated to make sense of the health gradient (Hertzman, 2009). Also other theories have contributed to this ongoing debate, such as social production model and eco-social theories (Krieger, 2001). Differences in hypotheses have helped shape public health strategies which, however, remain fundamentally political decisions and, as such, follow political cycles.

1.1.2.1.2   Behaviour change

How individuals make choices related to their health and how to influence related behavioural choices, is also a key issue for health promotion. Theories of health behaviour have been suggested from a range of disciplines including psychology, sociology, anthropology and economics (Davis et al., 2015). In conjunction with health equity theories, from arguments of patterning of illness due social factors to stances of social aetiology of disease several theses have been advanced as explanations for the health gradient (Cockerham, 2007).

Grounded in welfare economics, neoclassical theories have been proposed based on a perspective in which individuals are individualistic and rational calculating beings. As rational beings, individuals make decisions that are consistent with their aim of maximising their expected utility in the presence of uncertainty (Von Neumann J., 1944).

Notable among these theories, Grossman's model of health demand (Grossman, 1972) has seen widespread application in health economic research. Through this model, health is a source of utility that is inherited (i.e. health stock) and depreciates over time. Utility is generated both from avoidance of ill-health, as well as from availability of more sick-free days for leisure and producing income. Therefore, health is not only a consumer, but also an investment good (human capital). The rate of depreciation of the health stock depends on a range of factors including age, income and education, genetic, environmental and lifestyle factors.

The Grossman model is able to predict that socio-economically disadvantaged may be more likely to persist with unhealthy behaviours. Using income and education variables, a lower income means a lower return obtained from a healthy day and lower education means a lower efficiency in producing a given investment in health, compared to the more affluent groups. However, under a neoclassical framework, personal behaviour is deemed as the primary factor, with individual responsibility determining the health status, instead of socio-economic conditions (Henderson, 2014).

Since its publication, Grossman's original model has been extended and adapted (Wagstaff, 1986), for instance, with the incorporation of uncertainty into the model (Clark and Etile, 2002). Despite its popularity, however, this model has seen few applications to research on behaviour change (Davis et al., 2015).

Recently, van Kippersluis and Galama (2014) tested Grossman's model for explaining why wealthier individuals engage in healthier behaviours. These authors concluded that the richer individuals demand less unhealthy consumption as the health costs to them are greater than those to low income individuals. In his analysis, McCarthy (2006) evaluated the suitability of this model to predict decisions on whether to start health capital investments. He hypothesised that an increase in profit above what is rationally required would be needed before a regime, such as regular exercise, is initiated.

14

The assumption of rationality has been challenged by a growing body of research conducted in experimental and behavioural economics (The Behavioural Insights Team, 2010). A number of models have been proposed which relax some of the assumptions related to perfect rationality. Under a philosophy of libertarian paternalism, the concept of "nudge" has been recently introduced by Thaler (2008) gaining traction among researchers and enticing governments in the UK and US.

A conceptual framework (MINDSPACE), which collects a number of existing theories to change behaviour, has been recently proposed (Dolan, 2010). This framework is a summary categorisation of primarily automatic and contextual effects on behaviour that have been assessed in experimental settings (Vlaev et al., 2016). While the elements included in MINDSPACE represent a set of tools to influence behaviour, ethical concerns may arise with their use. In particular, questions may arise regarding who decides their structure and on what basis. People may not like the idea of government intruding into areas of personal responsibility. So, before decision-makers consider how they can apply these tools, it is important to guarantee that the public' view is represented.

Another pragmatic, but simpler framework developed by this team is called EAST (easy, attractive, social, timely, The Behavioural Insights Team, 2012). To date, no formal test of these frameworks has been performed. However, programmes based on theories in NUDGE and MINDSPACE, such as tobacco consumptions with interventions aimed at reducing smoking in

pregnancy and encouraging people to stop smoking, have been implemented (The Behavioural Insights Team, 2010).

A recent review of behaviour and behaviour change models found that just three theories accounted for over half of published applications (Davis et al., 2015). All of these focused on psychosocial factors and were, in order of frequency of application, "the Transtheoretical Model" (Prochaska and Velicer, 1997), "Theory of Planned Behaviour" (Ajzen, 1991) and "Social Cognitive Theory" (Bandura, 1985).

Focussing on the most widely applied, the Transtheoretical Model recognises behaviour change as a process that unfolds over time. Also referred to as "Stages of Change", this model conceptualises the process of intentional behaviour change, by breaking down behaviour change stages. In their paper, Adams J. and White M. (2004) challenged the validity of this framework in the context of physical activity promotion. These authors argued that this model is not superior to non-staged approaches in changing long-term behaviour, due the complexity of behaviour, the staging algorithm structure and the lack of inclusion of essential determinants of behaviour change. In response to this article, Brug et al. (2005) reviewed the evidence presented by Adams and White (2004) and found this model more likely to induce change in motivation and short-term behaviour change. Although the validity of the Transtheoretical model has not been established yet for complex interventions, this framework can be used as a way to conceptualise how in the context of population-level initiatives people at different states of "readiness" to change can react differently to exposure of an intervention.

1.1.3 Challenges in designing health promotion activities

While the application of theories has been advocated in intervention design and evaluation (Craig et al., 2008), in practice, health promotion initiatives are often launched for policy, rather than research purposes. These interventions typically emerge from past practice, policy makers and practitioners, and are implemented within several constraints to public health organisations (Davis et al., 2015).

As complex interventions, health promotion initiatives present a number of special problems that are in addition to the practical and methodological issues that evaluators usually face (Craig et al., 2008). To support researchers and decision-makers to recognise and adopt appropriate methods, the MRC published a methodolgical guidance on developing and evaluating complex interventions (Craig et al., 2008).

Being pragmatic, rather than theoretically grounded, is not the only aspect differing in practice from clinical / health care interventions. Health promotion intervention evaluations cannot often be designed as experiments, with individuals being recruited and followed up on an individual basis (Nutbeam, 1998). Rather, their evaluation typically relies on observational data of population samples. They are commonly non-research led, with primary data being collected and handled by non-research personnel and with analysts being involved only retrospectively, following large-scale implementation (Medical Research Council, 2010).

In addition, the cost of health promotion programmes is typically lower than clinical interventions (NHS Scotland, 2016) and they impact on multiple public budgets, involving decision makers outside the health care sector (Weatherly et al., 2009). In particular, the resources needed for and impacts generated from implementation of these initiatives can extend beyond health, with costs and benefits falling across different public agencies and sectors. This multi-sector approach presents challenges, both for design of intervention and evaluation of any potential benefits and costs. Examples of multiple-sector health promotion activities include public parks, school-based (Mikkelsen et al., 2014) or public policies to incentivise participation in sports activities (Allison et al., 2017) and mass-media campaigns around healthy eating (Espino et al., 2015).

1.1.4 Universal programmes

Universal approaches apply to entire populations, such as all residents, all employees in a workplace or pupils in a school, with access being based simply on being part of the defined population (Carey et al., 2015).

Universal programmes may advantage people in the population who are easier to reach and perhaps less or not in need of the intervention, thereby widening baseline health gaps (Niederdeppe et al., 2008). This may be especially true when interventions rely on voluntary behaviour change (Mechanic, 2002). In the context of healthy behaviours, being not in need of the intervention is meant as being already physically active or, using an example of dietary behaviours, eating the recommended daily portions of fruit and vegetables.

Conversely, targeted interventions apply to priority sub-groups within the defined population, which are determined by selection criteria. These criteria are often based on health equity agendas that are aimed to narrow existing health gaps across population sub-groups, which are deemed unfair and unjust. Both universal and targeted approaches present their own challenges.

Different forms have been combined so to maximise the strengths of each (Carey et al., 2015). Fundamentally universal in their health-promotive actions, blended approaches include targeted (Skocpol, 1991) and proportionate universalism (Marmot, 2010). An example of the former approach can be represented by a universal flu vaccine programme that includes an outreach strategy for high-risk groups. With a proportionate universal approach, instead, the scale and level of action is proportionate to the level of disadvantage. An example of this approach can be the provision of improved access to open green spaces (i.e. public parks) located in most deprived neighbourhoods. In this case, access is made available to everybody, though with geographical / logistical proximity representing the "proportionating factor".

1.1.5 *Physical activity and health promotion*

In countries that have passed through the epidemiological transition mentioned above, much of the public health policy focus has moved to promoting healthy lifestyles. A number of different categories of behaviours can fall under this broad umbrella term. Examples include drug and substance abuse (e.g. tobacco, alcohol and other psychoactive substances), unbalanced diets (e.g. overconsumption of sugar, salt, saturated and trans-

fatty acids, or insufficient intake of vitamins and fibres from fruit and vegetables) and physical inactivity (World Health Organization, 2016).

Lack of adequate and regular physical activity (PA) has been identified as a main cause of chronic disease and mortality in developed countries (Lee et al., 2012). Against this backdrop, international and national recommendations on PA have been issued. In the UK, the Chief Medical Office provides guidance on how much PA people should do (Department of Health, 2011a, Department of Health, 2011b). PA can be classified into four domains: occupational, home, transportation and leisure-time (Office of Disease Prevention and Health Promotion, 2018).

Current recommendations for adults are to engage in at least 30 minutes of at least moderate PA for five days a week. Moderate PA refers to a level of intensity of physical exertion, which requires a noticeable acceleration of the heart rate. Metabolic Equivalent Tasks (METs) are commonly used to express intensity of PA. MET is a measure of energy expenditure, which is represented by a ratio between the metabolic rate of an activity and the resting metabolic rate.

1.1.6 *Methods guidance for the economic evaluation of health promotion*

There exists relevant guidance to support analysts in evaluating complex interventions (Craig et al., 2008), behaviour change (National Institute for Health and Care Excellence, 2014, National Institute for Health and Care Excellence, 2007) and public health interventions (National Institute for

Health and Care Excellence, 2012). These recommendations cover a wide range of topics related to the development, implementation and evaluation of these interventions.

Acknowledging the practical constraints in which these programmes are often evaluated, Craig et al. (2008) recognised that, in these settings, researchers may have little or no say over how the intervention is implemented (i.e. non-research led). Given the cost of such interventions and likely fit in everyday practice (i.e. pragmatic), they recommend that evaluation should be carried out using the best available methods, even if they are not theoretically optimum.

Among other aspects, the methods guidance places particular emphasis on the importance of considering the issue of selection bias in observational data settings. Although randomisation remains the optimal option, this documents suggests a range of post hoc approaches, including the use of regression models and extensions, such as propensity score methods. While the choice of method is dependent on the analysis context and data available, it is recommended that analysts attempt to characterise the uncertainty related to the evaluative process inherent in the economic results to fully inform policymakers (Briggs and Gray, 1999).

Recommendations for evaluation of universal programmes to promote healthy behaviours are also dedicated to the incorporation of concerns regarding not only initiation, but crucially maintenance of behaviour change

over time into the EE (National Institute for Health and Care Excellence, 2007, National Institute for Health and Care Excellence, 2014). Emphasis is put on the monitoring of behavioural outcomes from short to the longer term, as well as on the heterogeneity of impact of these interventions by population sub-groups and on existing health inequalities.

Given the broader scope of guidance provision and variety of data settings that analysts can face, details on the analytical methods to use are not included in those documents. However, the use of decision analytic modelling has been supported by National Institute for Health and Care Excellence (2018), as a framework for meeting the requirement of an appropriate characterisation of decision uncertainty.

Compared to previous public health methods guidance, a The MRC framework and NICE public health methods guidance provide establishing the cost-effectiveness of health promotion activities, (National Institute for Health and Care Excellence, 2012). Specifically, a public sector perspective that needs to be able to reflect decision makers' interests and settings, beyond that of the health care sector. Recognising the imminent shift in responsibility (UK Government, 2012), from the NHS to Local Authorities (April 2013), it has been recommended that EEs should reflect both this wider remit by Local Authorities and the greater "local element" (National Institute for Health and Care Excellence, 2012). This change in approach means to assess the value for money of the intervention using a broader range of EE frameworks, other than CUA, and pay careful consideration on the specific decision-making contexts. To this respect,

recommendation is particularly concerned with the consideration of perspectives from the public department or local government that administers the intervention.

1.1.7 *Outstanding methodological challenges in economic evaluation of health promotion*

As discussed above, the MRC framework and NICE public health methods guidance provide general principles and advice on how evaluations and economic evaluations should be ideally conducted. In practice, however, the normative perspectives and practical constraints discussed in sections 1.1.2 and 1.1.3 often make the evaluation of universal programmes to promote healthy behaviour a difficult task. Indeed, reviews have showed that these added complexities have been addressed only in part in applied studies, with analysts often adopting methodological simplifications to enable an EE (Alayli-Goebbels et al., 2014, Bojke et al., 2018, Edwards et al., 2013, Hill et al., 2017, Owen et al., 2012, Weatherly et al., 2009).

The published literature is limited, often scattered and included in the so-called "grey-literature" (Population Health Science Research Network, 2012). However, examples of EEs of these programmes have been considered for the development of National Institute for Health and Care Excellence (NICE) recommendations on the design and implementation of interventions related to a variety of topics. For example, these include intervention modalities such as community engagement (National Institute for Health and Care

Excellence, 2016) and school-based settings (National Institute for Health and Care Excellence, 2010).

Focusing on methodological challenges that are not exclusive to, but typify these settings, estimating the effectiveness of health promotion activities may be complex. This is not least due to the level of missing data anticipated in such studies, which rely on data from population samples, but also due to the observational nature of most data generated (Craig et al., 2008). Recent reviews have noted low uptake of missing data methods (Leurent et al., 2018a), with simplistic, and often unrealistic assumptions regarding the mechanisms of survey non-response being made. The plausibility of a data collection process to obtain a sample of data that represents a random draw from the population may therefore be challenged.

Three aspects of such programmes are key for the modelling of treatment effect (i.e. intervention impact) for use in an EE in public health (Squires et al., 2016b). Firstly, intervention effects on disease risk cannot usually be fully observed within the programme duration. Therefore, extrapolation of effects over longer time periods is required.

Secondly, non-clinical populations are typically the target of these initiatives. As (apparently) healthy individuals, a range of chronic diseases is considered. Occurrence of diseases is dependent on lifestyles, as well as on personal characteristics (e.g. age or sex-related conditions, such as dementia and prostate cancer) and changes in healthy behaviours can lead

to overall reductions in disease risk. However, disease occurrence

probabilities may not be independent from one another. For instance, risk of

type II diabetes has been associated with risk of colorectal and breast

cancers, mainly due to shared risk factors, among which health behaviours

play a major role (Giovannucci et al., 2010).

A third aspect to consider is that of population-level impact. In particular, the

heterogeneity of impact that these intervention can have on different

individuals in the population. In these settings, standard one-to-one

comparisons between "average" individuals may not be adequate (Squires et

al., 2016b).

In terms of economic analysis, methodological reviews have also noted that

the majority of EEs have been conducted to inform decisions from a health

care sector / NHS perspective (Alayli-Goebbels et al., 2014, Weatherly et al.,

2009). In particular, studies have focussed mostly on the impact of the

intervention on health care budgets, with the incorporation of equity

concerns, which represents a primary objective in public health especially

with universal approaches, being mostly unaddressed. In some instances,

the multi-agency nature of these programmes and shorter financial cycles of

public sector entities have motivated some researchers to explore alternative

and broader perspectives, such as Doring et al. (2018) and Frew et al.

(2014).

1.2 Thesis aims and objectives

The methodological challenges discussed above show that, in establishing the cost-effectiveness of health promotion activities, there are a number of areas that may require additional efforts. Previous methodological reviews have showed that choice of analytical methods to assess the cost-effectiveness of public health interventions, especially universal programmes to promote healthy behaviours, has often been pragmatic and key methodological issues not being addressed (Alayli-Goebbels et al., 2014, Weatherly et al., 2009). These methodological issues are: modelling of heterogeneity and inequality in population-level impact, assumptions regarding maintenance of behaviour change over time, longitudinal selection bias and perspective for economic evaluation.

In order to explore these issues, the application was narrowed to the field of PA promotion. This is an area in which there has been very little consideration of the appropriateness of methodological approaches, specifically those available more generally for public health.

Three research questions were therefore formulated:
1. If and how have the above acknowledged methodological challenges have been addressed in EEs of PA promotion interventions?
2. How to incorporate concerns regarding population-level impacts into the economic model;

3. What is the impact of the key methodological assumptions that underpin the existing models on the economic decision?

A mixed method approach was used to achieve these objectives. In order to identify relevant shortcomings of existing economic models, a systematic review and critique of the EE literature were undertaken. A case study provided an example of the limitations related to data collection process and decision-making context. The case study also provided primary data that were analysed to populate a decision-analytic model, which was developed to address the identified modelling shortcomings. Case study cost-effectiveness results were subsequently used as a benchmark to test relevant modelling and methodological assumptions and illustrate their implications on the cost-effectiveness decision.

1.3   Structure of the thesis

The thesis is structured into 7 chapters. Chapter 2 reviews the existing EEs of PA promotion. Chapters 3 and 4, which are titled according to the first two phases of EE respectively (i.e. estimation of effectiveness and modelling of impact), include both methodological work (i.e. addressing of the identified challenges) and applied work (i.e. analysis of the case study).

Chapter 5 focuses on the EE of the case study. Chapter 6 first focuses on methodological aspects relating to the economic analysis of universal programmes of health promotion, to then move onto testing key modelling and methodological assumptions that underpinned previous economic

models. Chapter 7 provides a discussion of the thesis findings and limitations, concluding with a summary of the lessons learnt and identifying areas for further research. Below a description of the chapter contents is provided.

Chapter 2 - Literature review includes the methods and results of a systematic review of EEs of PA promotion. An overview of the study findings and critique of the analytical methods used to assess cost-effectiveness is provided, with a focus on the issues related to the evaluation of universal programmes. Based on the review findings, the chapter concludes with a section which identifies four methodological issues to address (chapter 3, 4 and 6) and four sets of modelling and methodological assumptions to test (second part of chapter 6).

Chapter 3 - Estimation of effectiveness is centred on the econometric analysis of the data. This chapter is split into three parts. It first describes the case study decision-making context, programme contents and data collection process. A description of participants' characteristics is then followed by an estimation of effectiveness parameters to populate the developed decision-analytic model (chapter 4), and subsequently perform the EE of the case study (chapter 5). The focus then moves onto the issue of selection bias. This is addressed formally, by specification of three alternative scenarios regarding the causes of second-stage survey non-response, for further testing of cost-effectiveness results to conduct in chapter 6.

Chapter 4 - Modelling of impact presents a modelling approach devised to overcome the two identified modelling shortcomings of previous models (i.e. population-level impact and behaviour change maintenance over time). The first part is dedicated to describe this approach and its mechanics. The second part describes a decision-analytic model, which was developed based on the proposed modelling approach and designed to assess the value for money of universal programmes to promote PA. Sub-sections are dedicate to describe how the decision-analytic model's general structure was adapted to enable an EE of the case study.

Chapter 5 - Economic evaluation of the LLGA case study presents the results of a cost-effectiveness assessment obtained using an evaluative approach consistent with previous EEs. Using the decision-analytic model presented in chapter 4, base-case cost-effectiveness results are presented.

Chapter 6 – Exploring the implications of modelling and methodological assumptions explores the impact of variations to the four groups of assumptions identified at the end of chapter 2, using the cost-effectiveness results obtained in the previous chapter as the benchmark. These assumptions concern: selection bias, range / combination of modelled diseases, behaviour change maintenance over time and perspective for EE. Before proceeding with the analysis, the methods section describes how a change in perspective is addressed. Estimates of deviation from base case cost-effectiveness results, their graphical illustrations and implications for decision-making are then represented.

Chapter 7 - Discussion provides a critique of the significance and limitations of this work, focussing on the thesis findings and placing them within the relevant literature. The chapter concludes with a summary of the lessons learnt that could be useful for future evaluations and identifies areas for further research.

2 Chapter

Literature review

2.1 Chapter outline

This chapter presents the findings of a systematic review conducted to critique the EEs of PA. Before the systematic review was started, a review of existing reviews of EEs of PA was performed. This was to outline how existing reviews had already investigated the subject area, providing a rationale and justification for the systematic review. The systematic review aimed to answer the following research question:

- How the methodological challenges, relating to attribution of effects, measuring and valuing outcomes, inter-sectoral costs and consequences and equity considerations that have been highlighted in previous reviews of public health interventions, have been handled in practice in EEs of PA promotion interventions.

The last section of this chapter identifies the outstanding methodological issues in the existing applied literature that will be addressed in the thesis.

2.2 Search strategy

The search strategy was developed following recommendations by relevant guidelines (Centre for Reviews and Dissemination, 2009, Joanna Briggs Institute, 2017). The search for relevant papers for the two reviews (i.e. reviews of existing reviews and review of EEs) was based on the same search strategy. Different eligibility criteria, study screening and selection

31

process were applied for the two reviews. Details on the latter are provided in the respective review sub sections.

2.2.1 Development

Given the interdisciplinary nature of the investigated topic, relevant studies could be found both in specialized and generic databases. Searches were conducted using five electronic databases, which were chosen according to their potential coverage. In order to achieve comprehensiveness of the review, databases were selected according to their likelihood to retrieve from them, more specifically from their indexed journals, most of the relevant evidence available. However, a certain degree of overlap between databases, in terms of number of duplicates, was expected.

A search concept tool as applied to structure the inclusion criteria. A method consisting of using some of suggested "PICOS" concepts (i.e. "I" for intervention and "S" for type of study") was chosen (Centre for Reviews and Dissemination, 2009). Free-text terms, synonyms, spelling variants, abbreviations and indexing terms (e.g. subject headings) related to three concepts were used: (1) EE, (2) economic model, (3) PA. No manual search using reference lists of existing literature was planned. Appendix M reports the search strategy.

Validated search filters for identification of the relevant literature were not available. Search strings were developed from terms identified in known relevant publications and related to those three concepts. Concepts were combined using Boolean logic, as follows: (1) EE "OR" (2) economic model

and the resulting (1+2) "AND" (3) PA. Other search filters, such as for intervention setting or type (e.g. community-based or workplace), were not included as eligible papers could be missed. No limit to publication date or to the unpublished literature were set. Although an undoubtedly high yield of studies was expected using this approach, high sensitivity of the search was guaranteed.

2.2.2 Testing

In order to obtain a certain degree of precision while keeping a balance in terms of sensitivity, the search strategy was tested for level of specificity (Brettle et al., 1998). There is no standardised approach for this testing (van Mastrigt et al., 2016). Following methods comparable to those used in other known peer-reviewed reviews (Alayli-Goebbels et al., 2014), this was undertaken by screening the titles of a random 10% of the total number of records.

Given the high proportion of non-relevant papers found at this stage (>95%), the search strategy was rerun focusing some of the subject headings used. With this additional step, the previous testing was repeated, but the results were still deemed not acceptable as they yielded a low proportion (10%) of potentially relevant articles. Thus, a forth concept of (4) behaviour/lifestyle was added to the search strategy so as to make it more precise. The concept (3) PA was combined "OR" (4) behaviour/lifestyle, for the resulting combination of (3+4) "AND" (1+2).

There is no established or agreed proportion of non-relevant papers retrieved for the strategy to be regarded as correct, as it is dependent on the type of search, topic and aim of the review (van Mastrigt et al., 2016). By the addition of the forth concept, and after consultation with information specialists, balance between sensitivity and precision was deemed to be achieved. The strategy delivered a manageable amount of records (6,951 hits) and a valid proportion of potentially relevant references by title (20%). As a result, a successful strategy was obtained. Following guidance by Centre for Reviews and Dissemination (2009), the strategy was also tested on Medline as to whether it retrieved relevant reviews of EEs already known, but that had not been used to develop the draft strategy. Having obtained a positive test, the final strategy was thus ready for MEDLINE database. It was subsequently converted in the other four databases, where it was adapted to take into account of the differing search options, indexing terms and search operators across databases. All search strings were saved for each database.

2.3   Review of existing reviews

2.3.1   Methods of review

2.3.1.1   Eligibility criteria

Any reviews of EEs on the investigated topic were eligible for assessment. Economic analysis, such as those by Laine et al. (2014) and (Wu et al., 2011), o reviews of EEs focusing only on interventions promoting PA in combination with other healthy behaviours or technologies (e.g. dietary habits) were excluded.

2.3.1.2   Study screening and selection

All the retrieved references were managed using Endnote 7 software. A

library file was created for each of the five identified databases, so as to keep

track of records. After having merged the five Endnote libraries in one file,

removal of duplicates was performed following a University of Leeds Library

recommended guide (Academic Unit of Health Economics, 2014). Relevant

literature was then identified using EndNote keyword search tool. In

particular, the word "review" was searched by document title.

The identified papers were screened by title and abstract level for relevance.

For each identified review, information concerning scope and approaches to

evidence synthesis and quality assessment were extracted from the full texts.

2.3.2   Results

Six records fulfilled the inclusion criteria. Four focused on reviewing the

economic evidence for PA promotion interventions, whereas the remaining

two (Alayli-Goebbels et al., 2014, Gordon et al., 2007) also included EEs of

other healthy behaviour programmes (e.g. smoking cessation) as face-to-

face and behaviour change interventions, respectively. Five reviews focused

on the economic evidence for specified intervention modalities or settings,

while only one (Alayli-Goebbels et al., 2014) focused on appraising the

methods used for EE. The methodological review by Alayli-Goebbels et al.

(2014) updated on the progress made, within the behaviour change area,

with the handling of previously identified methodological challenges

characterising the EE of public health interventions (Weatherly et al., 2009).

2.3.2.1   Characteristics and contents of the identified reviews

Date of publication of the studies included in the reviews ranged from 1981

(Alayli-Goebbels et al., 2014) to the more recent review including EEs

published up to August 2014 (Gc et al., 2016). No limit to search was set

within the reviews in terms of country of study or investigated healthcare

system, although the English language was used as a search filter in the

majority of cases. Overall, the reviews assessed relatively low numbers of

studies, with a minimum of four (Pavey et al., 2011) to a maximum of 17

reports (Alayli-Goebbels et al., 2014, Gordon et al., 2007). This number

depended on the search strategies adopted (e.g. number and type of

databases searched) to retrieve references potentially relevant to the posed

research questions within the set scope, as well as on the time of review.

2.3.2.2   Scope of the reviews

All the six reviews considered economic evidence for adults, with three of

them (Garrett et al., 2011, Gc et al., 2016, Muller-Riemenschneider et al.,

2009) focussing the investigation on that age group. In two of the six reviews

(Garrett et al., 2011, Gordon et al., 2007) no restriction was set in terms of

health state of participants, who could be either healthy or at an increased

health risk.

The reviews by Gc et al. (2016) and Pavey et al. (2011) included only

interventions on inactive individuals, whereas Muller-Riemenschneider et al.

(2009) limited the inclusion to studies on healthy individuals. With regard to

promotion level, Garrett et al. (2011), Gc et al. (2016), Gordon et al. (2007)

and Pavey et al. (2011) assessed the economic evidence for PA initiatives targeting individuals (e.g. exercise referrals) or delivering the intervention at individual level (e.g. face-to-face). These reviews concentrated on assessing specific PA interventions, such as brief advice or counselling, with two of them (Garrett et al., 2011, Gc et al., 2016) limiting the investigation to defined promotion settings (i.e. primary care and the community).

The other two reviews (Alayli-Goebbels et al., 2014, Muller-Riemenschneider et al., 2009) did not limit the inclusion of studies to type of PA intervention. Alayli-Goebbels et al. (2014) focused on behaviour change initiatives. Except for Alayli-Goebbels et al. (2014) and Gc et al. (2016), cost per increase in outcome or cost-effectiveness data were used as minimum requirements for inclusion of a study. As for the accepted comparators, no requirements were instead set in the eight reviews.

The reviews by Pavey et al. (2011), Alayli-Goebbels et al. (2014) and Gc et al. (2016) did not limit the inclusion as to type of study design. Gordon et al. (2007) restricted the review to full EEs, Muller-Riemenschneider et al. (2009) excluded CUA and CBA designs. Garrett et al. (2011) included randomised controlled trial (RCT)-based EEs with at least 6 months of follow up intervention period and effectiveness trials recruiting a minimum of 50 participants, respectively.

2.3.2.3 Approaches to evidence synthesis and quality assessments

Of the five reviews focussing on the economic estimates, none used a meta-analytic approach to synthesize quantitatively the results of individual studies.

Four reviews (Garrett et al., 2011, Gc et al., 2016, Muller-Riemenschneider et al., 2009, Pavey et al., 2011) summarised narratively the economic results by ranking the included studies in order of incremental cost-effectiveness ratios (ICERs). One study (Gordon et al., 2007) described the ranges of ICERs reported (i.e. cost per QALY gained).

Disparate approaches to quality assessment and instruments were used across the reviews. The methodological review by Alayli-Goebbels et al. (2014) assessed quality in terms of progress made with regard to identified methodological challenges. For the other five reviews, instead, the focus was put broadly on assessing the validity of individual study results. Nonetheless, there was wide variation on how this was achieved.

Across reviews, the reviewers used standardised (or adapted from) quality checklists to assess risk of bias in the source of effectiveness evidence (Muller-Riemenschneider et al., 2009), level of reporting in EE (Alayli-Goebbels et al., 2014, Garrett et al., 2011, Gc et al., 2016, Gordon et al., 2007, Laine et al., 2014, Pavey et al., 2011) and compliance with good practice in decision-analytic modelling (Pavey et al., 2011). The respective risk/quality levels were used mostly to rank the primary studies, but they did not influence the inclusion of the articles nor were used for weighting the estimates.

Checklist scores were used as crude quality measures, being often accompanied by only brief discussions on the shortcomings of the reviewed studies, particularly on the limitations deriving from the design of

effectiveness studies (e.g. short follow-up periods, non-experimental designs) and the consequent low quality of available data.

Except for two reviews (Alayli-Goebbels et al., 2014, Gc et al., 2016), validity of the economic results was not appraised in light of the methodological assumptions underlying the economic analyses (e.g. life-cycle costs, long-term effectiveness). Gc et al. (2016) pointed out the exclusion from the analysis of relevant costs related to implementation of the intervention (out-of-pocket expenditures), whereas Gordon et al. (2007) assessed the assumptions made about compliance, relapse behaviours and long-term effectiveness. In addition, no review appraised critically the EE methods (e.g. perspective, time horizon) chosen for the analyses.

2.3.3 Conclusions

A marked heterogeneity was found in terms of review focuses and approaches, making direct comparison of study findings particularly difficult. Results indicated that the area of PA promotion was explored only in part by existing reviews of EEs. Across reviews, the majority of research efforts were directed towards summarising cost-effectiveness evidence for specific PA intervention modalities. Methodological assessment of the included studies was limited in scope, with only one review (Alayli-Goebbels et al., 2014) focussing on the progress made in regard to methodological challenges acknowledged in previous studies. A comprehensive and up to date systematic review of EEs of PA promotion interventions was therefore deemed necessary to shed light on the methodological issues and

assumptions underlying the economic studies, particularly those assessing

the cost-effectiveness of universal strategies.

2.4   Systematic review

2.4.1   Methods of review

The objectives were to:

1. To describe the contexts, analysis designs, and type of interventions;

The studies were grouped according to type of approach employed, following

a previous methodological review (Griffiths et al., 2012). Namely, trial-based

pure modelling or mixed-methods studies. In trial-based analyses, the

sample of participants and follow-up period of the effectiveness study match

those considered for EE. In mixed-methods studies, the EE is based on a

single effectiveness study, but extrapolation is performed modelling effects

over longer periods and/or to other contexts. In pure modelling studies,

effectiveness parameters are obtained from secondary sources (typically

meta-analyses of trials) and used to simulate hypothetical scenarios for

defined THs.

2. To summarise cost-effectiveness results and findings;

If reported in the full text, incremental ratios were grouped and presented as:

benefit/cost ratios, incremental costs per QALY or disability-adjusted life year

(DALY) gained and incremental costs per additional units of intermediate and

final outcomes. For the latter group of studies, a brief summary is given. To

compare the economic results between studies, values were all converted in

2017 equivalent £ sterling using a health specific inflation converter (EPPI, 2019), with midpoints presented for studies reporting ranges of results.

3. To critique the analytical methods used for EE.

No quantitative analysis (i.e. meta-analysis) of the study estimates was planned as deemed beyond scope. A critique of the methods used for EE of the included interventions was given in the form of narrative summaries. This method has been recommended for methodological reviews and when study heterogeneity is particularly pronounced (Joanna Briggs Institute, 2017). The idea of using of a quality appraisal checklist, such as CHEERS (Husereau at el. 2013), was discussed but discarded. This was because quality of reporting was not an aim of the review. However, this checklist and the framework used by the methodological reviews by Weatherly et al. (2009) and Alayli-Goebbels et al. (2014) informed the structure of the critique and the identification of the items/headings.

2.4.1.1 Eligibility criteria

Prospero database confirmed the absence of any ongoing reviews. Studies were eligible for inclusion if they met the following criteria:

- Type of study: any type of full EE. Partial EEs, such as cost-analyses were excluded.

- Intervention: any intervention aimed to promote PA behaviour (being either the focus of the study or one of the comparator interventions). Curative or rehabilitation programmes or studies evaluating the impact of hypothetical scenarios of changes in behavioural patterns (e.g. shift

in number of active travellers) or associated health risks were excluded. As were those promoting PA in combination with other technologies or interventions (e.g. health dietary habits). Combined interventions cannot be fully comparable, because they address different yet closely related and multifaceted issues (e.g. obesity) and it can be particularly difficult to disentangle the combined effects on the economic results.

- Population: non-clinical populations. EEs whose study populations were targeted or selected on the basis of pre-existing disease conditions were not included (i.e. disabled individuals or secondary interventions in cardiac patients). Studies targeting "high risk" individuals, that is, clinically stable but carrying medically relevant conditions, such as hypertension or mild/moderate depression were included.

- Written in the English language (to allow for cross-checking).

2.4.1.2 Study screening and selection

Identification of relevant articles was performed by screening against inclusion and exclusion criteria. If there was insufficient information in the retrieved article, the corresponding author/s were contacted to obtain the full text. After removal of duplicates, initial screening of titles against inclusion criteria was undertaken. This step resulted in a number of records to screen

by abstract, with excluded references that were grouped in relevant categories.

Screening of abstracts followed, excluding articles on the basis of study type (i.e. not full EE), intervention type (i.e. not solely on PA promotion) and target population (e.g. cardiac patients). Following a procedure comparable to that followed in the review by Alayli-Goebbels et al. (2014), a random 20 percent of the articles screened by title and abstract and all of the records assessed full text were reviewed by a second researcher. Any disagreement was resolved through discussion.

2.4.1.3   Data extraction

Data extraction forms were developed by adapting existing templates suggested by review guides (Centre for Reviews and Dissemination, 2009, Joanna Briggs Institute, 2017) and in the reviews by Weatherly et al. (2009) and Alayli-Goebbels et al. (2014). These forms were designed to capture contextual and key methodological elements relevant to the set objectives (Centre for Reviews and Dissemination, 2009). For all studies, only the information presented in the original publication was used.

2.4.2   Overview of the included studies

The systematic search yielded a total of 6951 records. After removal of duplicates, articles were screened by title. The majority of articles was discarded at this stage as lacking of requirements for inclusion (e.g. non-EE studies). After screening the abstracts, 54 full texts were selected for retrieval. Two articles referring to primary papers published by different

authors (Cavill, 2011, Petrella, 2006) were retrieved in full text and included, while 16 were dropped as failing to meet inclusion criteria (e.g. partial EEs). Thirty-eight unique articles fulfilling the inclusion criteria were thus included. A PRISMA-style diagram depicting the flow of information through the different phases of identification, screening and selection is displayed in appendix A.

The included studies were grouped by relevant categories to allow for discussion of their main characteristics. They are summarized in Tables B.1 to B.5 (see appendix B). A large degree of heterogeneity between studies was found, making it difficult to synthesise them into coherent messages. However, an outline of the included EEs is provided in the following paragraphs.

2.4.2.1 Contexts

The majority of studies (20/38) were published in the previous six years, confirming a marked upward trend in the number of EEs performed in the research area, which more than doubled compared to the previous two decades (1990-2010). However, this growth in economic research was not spread evenly across countries. Two-thirds (25/38) of the empirical investigations were conducted in or concerned the health systems of the UK or the USA. Five studies were carried out in the Australian continent and continental Europe contexts (The Netherlands, Spain and Belgium) each, only one in Asia (Taiwan), one in Canada, and the remaining project across multiple countries, namely, Mexico, Colombia and USA (Montes et al., 2012).

2.4.2.2   Analysis designs

Figures 2.1, 2.2 and 2.3 below show the proportions of studies for each of the three analysis approaches and break them down by follow-up period length. Given that gains in health outcomes are dependent on how the intervention effect is sustained over time, it was important to highlight the follow-up period of the effectiveness studies used to inform the economic evaluation and whether modellers applied extrapolation methods to represent future trends in PA changes over time.

Figure 2.1 Trial-based analyses
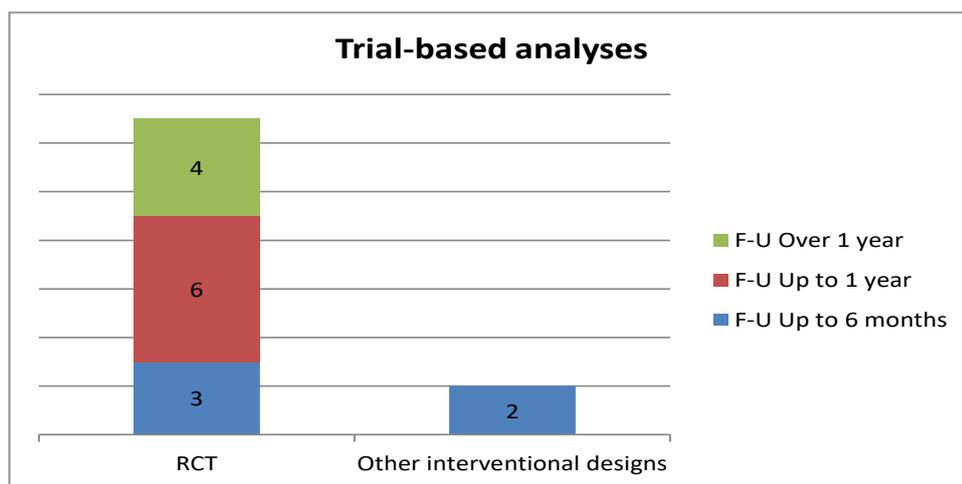
Note: F-U = follow-up period

The large majority of studies (13/15) was based on RCT designs, with only four studies (Groessl et al., 2016, Munro et al., 2004, Sevick et al., 2000) following up participants for 2 to 2.6 years. Two studies employed other types of interventional designs, more specifically, one controlled (Chen et al., 2008) and one single-arm trial (Vestergaard et al., 2006) with short follow-up periods (i.e. < 6 months).

Figure 2.2 Mixed-methods studies
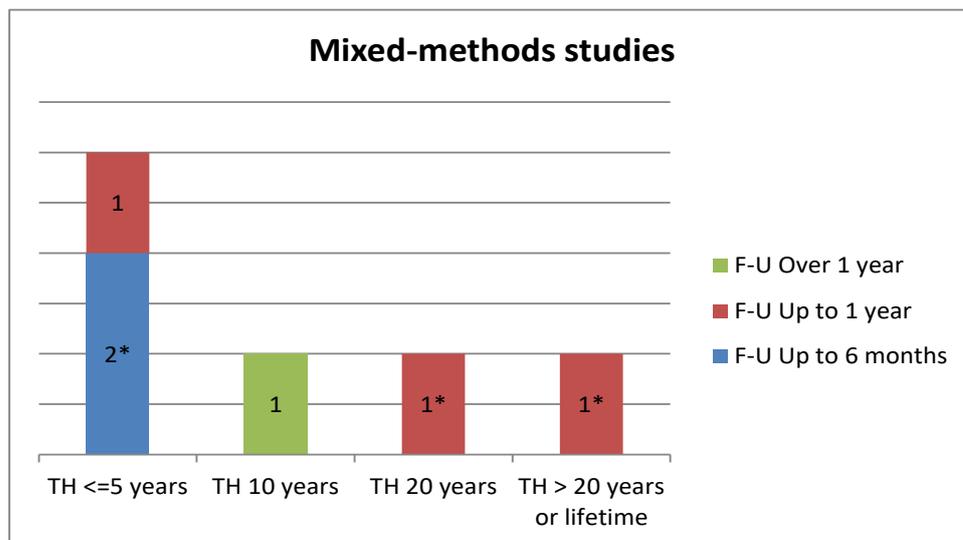
*extrapolation of results to the whole target population

Six studies employed mixed-methods approaches, of which (Cavill, 2011)

only one had a follow-up period longer than one year (i.e. three years,).

Except for De Smedt et al. (2012), who employed a RCT design,

effectiveness was measured through analysis of data from single-arm

studies. Four studies assessed incremental costs and consequences for

scale-up intervention scenarios (De Smedt et al., 2012, Frew et al., 2014,

Moodie et al., 2009, Peterson et al., 2008). Long-time horizons (> 5 years)

were considered in three studies (Cavill, 2011, De Smedt et al., 2012,

Moodie et al., 2009), of which only one study extrapolated results over a

lifetime period (Moodie et al., 2009).

Figure 2.3 Pure modelling studies
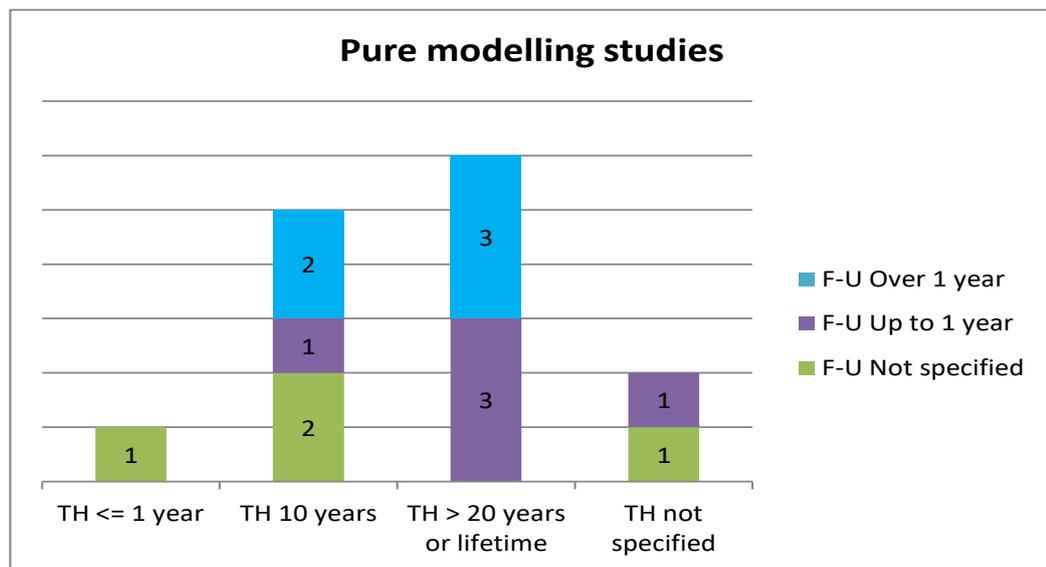
Of the 14 modelling studies, five were based on observational evidence (Amarasinghe, 2010, Dallat et al., 2014, Guo and Gandavarapu, 2010, Over et al., 2012, Wang, 2005), such as census reports or surveys, while the remaining eight studies on reviews or meta-analyses of trials. Only two studies (Anokye et al., 2014, Gulliford et al., 2014) set a 12-month minimum of follow-up period for inclusion of a study. Only five studies included evidence from follow-up periods longer than one year. Babey et al. (2014) measured effectiveness by reviewing and analysing evidence from identified interventional studies, but they did not disclose any follow-up criteria for inclusion of a trial.

Except for two studies using prevalence data (Wang, 2005), modelling studies considered long time horizons, with six of them (Anokye et al., 2014, Cobiac et al., 2009, Dallat et al., 2014, Over et al., 2012, Roux et al., 2015,

Roux et al., 2008) modelling the health economic impact of interventions over lifetime periods.

The remaining three studies (Beale et al., 2012, Goyder et al., 2014, Pringle et al., 2010) used a combination of approaches for the primary analysis, where both modelling techniques and within-trial approaches were performed.

2.4.2.3   Types of intervention

Twenty-one of the 38 articles included an analysis or a simulation of the impact of a universal initiative to promote PA behaviour. Target populations were general populations of adults in the majority of cases (33/38), one analysis focused on adolescents (Peterson et al., 2008) and four on children (Babey et al., 2014, Barrett et al., 2015, Moodie et al., 2009, Sutherland et al., 2016). In implementation terms, the four analysis focusing on children (6-12 years) were based on interventions in school settings (e.g. physical education, active commuting to school), McEachan et al. (2011) analysed a workplace intervention, while the remaining papers focused on community-based programmes.

The programme analysed by Frew et al. (2014) employed a universal approach to PA promotion. The remaining analyses focused on interventions targeting inactive adults from the general population (9/17) or at "higher health risk" (based on ethnicity, medical conditions, or areas of residence).

Using the classification proposed by Michie et al. (2011), universal programmes were multicomponent in the majority of cases. Three studies analysed communication strategies (e.g. mass media campaigns, facilitators), four changes to the physical environment (e.g. sidewalks, bike trails) and six service provision policies (e.g. exercise classes). The large majority of the analysed targeted intervention focused on communication or service provision policies. Nshimyumukiza et al. (2013) simulated the impact of a mass-media campaign, while Beale et al. (2012) and Dallat et al. (2014) the impact of changes to the physical environment, all on hypothetical cohorts of inactive adults.

2.4.3 *Economic evaluation results*

Figure 2.4 and 2.5 summarise the benefit/cost and incremental cost-effectiveness ratios presented in the reviewed studies.

Figure 2.4 shows that all CBAs reporting cost-benefit ratios found PA interventions providing positive return of investments. All these five analyses were based on changes to the physical environment.

Figure 2.4 Benefit/cost ratios

Benefit/cost ratios

Study	£ in return per £ 1.00 invested
Building multi-use and cycle infrastructures (Beale et al., 2012)	11
Community-based multicomponent programme (Cope et al., 2010)	2.44
Building sidewalks and bike trails (Wang et al., 2005)	2.35
Community-based multicomponent programme (Montes et al., 2011)	2.11
Building sidewalks (Guo and Gandavarapu, 2010)	1.5
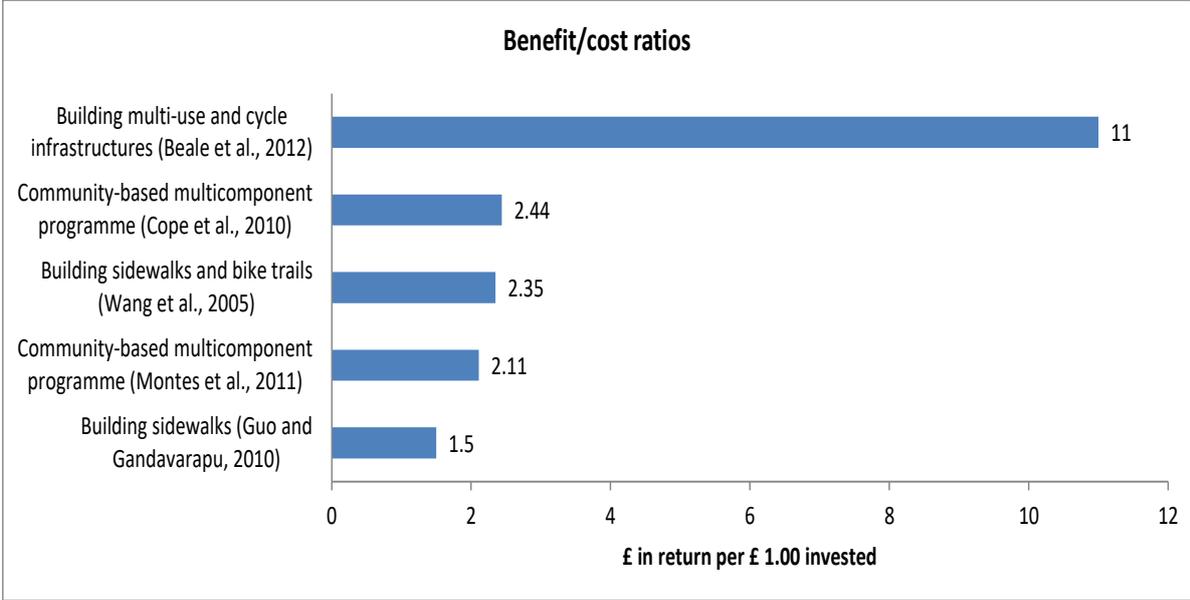
Beale et al. (2012), who simulated the impact of a targeted strategy on inactive adults (first line), reported the highest return on investment with estimated £11 per £1 invested, compared to the remaining four universal programmes. The lowest return on investment was estimated at £1.5 in the study by Guo and Gandavarapu (2010) who assessed the value for money of building walkways to promote active commuting.

Alongside a CUA, Frew et al. (2014) conducted a willingness-to-pay exercise based on a contingent valuation methodology. The monetary value of the programme from participants was elicited, at baseline and follow-up and a positive net benefit equal to £96 per participant was estimated. By using a willingness-to-pay approach, health and non-health benefits related to the programme and perceived by participants were included in the evaluation. This overcomes the limitations of using a simple benefit/cost ratio approach,

however the elicitation methods used in that analysis have been found to be prone to range bias (Whynes et al., 2003).

Furthermore, from a decision-making standpoint, cost-benefit and return on investment ratios may provide easy-to use decision-support tools (Pokhrel, 2015; Masters et al., 2017). Recently, a CBA-based model has been developed to support Local Authorities in England in the development of business cases for new and innovative interventions (Holden and Harding, 2015). Specifically, this model is rooted in a social return on investment approach. It follows what is recommended within the Green Book five case model (HM Treasury, 2018), and enables the wider economic value of public health interventions, including social benefits in terms of improved health and well-being, to be captured.

Figure 2.5 below ranks the studies by incremental cost per QALY (in blue) or DALY (in red). A wide variation in magnitude of incremental cost per QALY or DALY was shown. Fourteen of the 17 incremental ratios were below the respective commonly accepted willingness-to-pay thresholds, as applied in study countries.

Figure 2.5 Cost-utility analyses: incremental costs per QALY and DALY
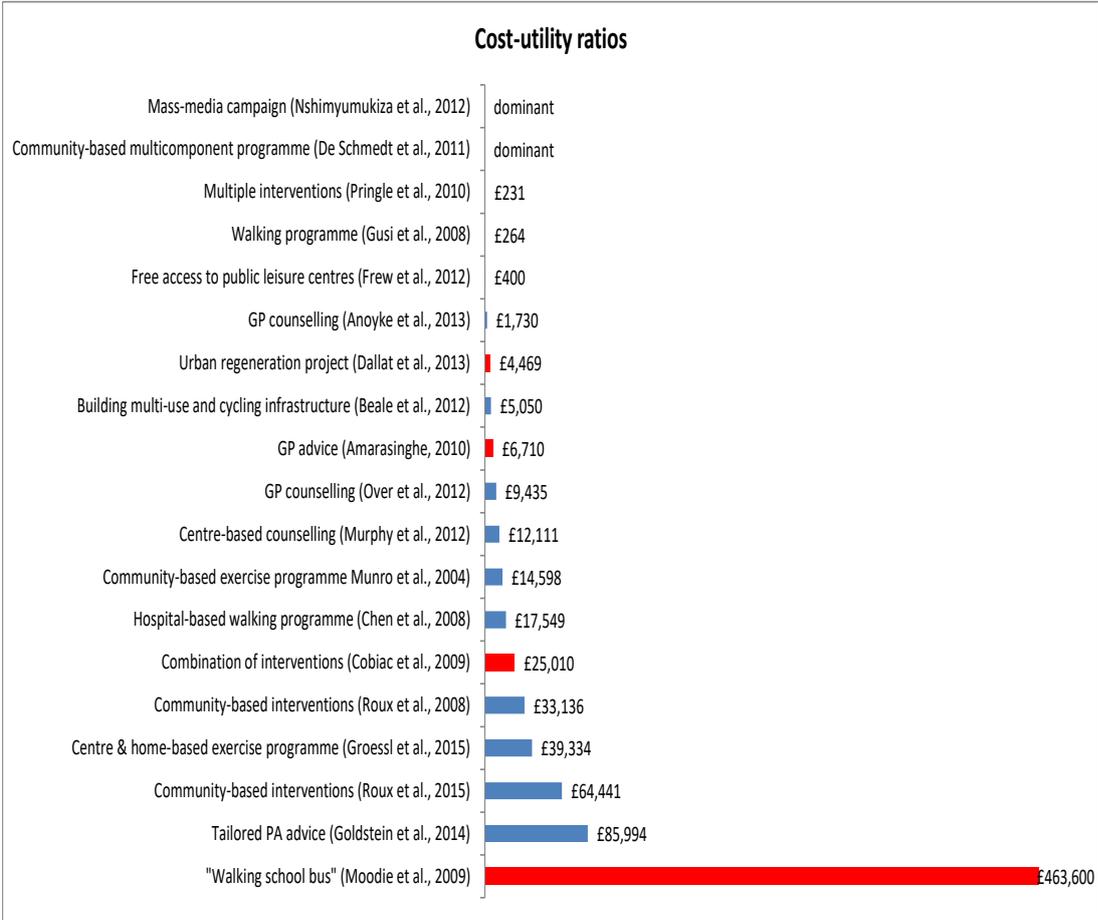
For two of the 21 CUAs (Gulliford et al., 2014, McEachan et al., 2011), no incremental costs per QALY or DALY were reported. Gulliford et al. (2014), through a probabilistic model, calculated the number of incremental QALYs (3.2 for 5 year and 5.0 for ten year time horizon) per 1000 participants entering intervention (valuing one QALY £30,000), while McEachan et al. (2011), who found the intervention to be not cost-effective, observed an incremental net monetary benefit (INMB) of -£103,02 (valuing one QALY £20,000). Both universal and targeted approaches varied widely in terms of CUA results.

As for the latter category of studies, five papers reported CEA on final outcomes, but did not express these as QALYs or DALYs. Groessl et al. (2016), who evaluated a centre and home-based exercise programme, estimated £33,901 per major mobility disability prevented. Nshimyumukiza et al. (2013) found promoting PA (the intervention was not described) as a dominant strategy (i.e. the intervention costs less and is at least as effective as the comparator) when estimating the number of osteoporosis-related fractures. A wide range of incremental results was instead found across the four studies estimating costs per LY saved (Munro et al., 2004, Over et al., 2012, Roux et al., 2015, Roux et al., 2008), with values ranging from £330 to £120,668.

A number of studies reported intermediate outcomes in CEA. Costs per MET-hour gained was estimated for four studies (three school-based programmes, (Babey et al., 2014, Barrett et al., 2015, Sutherland et al., 2016) and tailored PA advice (Golsteijn et al., 2014), ranging from less than £ 0.01 to £ 0.80. For studies estimating increases in PA minutes, such as PA counselling, (Larsen et al., 2015), behavioural skills and structured exercise programmes (Sevick et al., 2000), school-based activities (Sutherland et al., 2016), a range of incremental costs was found between £0.12 and £34.16.

Across four studies which assessed a mass-media campaign (Peterson et al., 2008), a range of community-based interventions (Pringle et al., 2010), a walking and counselling programme (Shaw et al., 2011) and a home-based exercise programme (Stevens et al., 1998), costs per person moving to a

more physically active category were calculated. However, how becoming more active was defined varied noticeably across the studies. The range of these results was broader than the former measures, ranging from £7.09 to £1,523. Only one study (Sutherland et al., 2016) calculated the cost per unit of body mass index (BMI) avoided, at an incremental cost of £1,126.

In the remaining four studies (Dallat et al., 2014, Goyder et al., 2014, Isaacs et al., 2007, McEachan et al., 2011), other measures of intermediate / surrogate outcomes (e.g. energy expenditure and clinical biomarkers) and health measures (e.g. mortality rates and non-preference based measures) were used to compare the alternatives.

2.4.3.1   Interpretation of the study results

According to the incremental ratios and estimates presented above and reported in the included studies, the evaluated PA promotion interventions were likely to be cost-effective and generally considered good value for money alternatives. However, no clear patterns could be identified. In particular, no intervention modality or setting was found superior to others in terms of cost-effectiveness. Nevertheless, according to the reported estimates, interventions involving changes in the physical environment or providing exercise opportunities to adults appeared to provide better value for money compared to other types of PA interventions. These programmes were in the majority of cases universal initiatives.

2.4.3.2   Uncertainty around study results

As acknowledged by the study authors, although to differing extents, results

estimated in the EEs were associated with various degrees of uncertainty.

The majority of studies (30/38) explored some form of uncertainty analysis. In

particular, the results were tested for sensitivity of variations in input data,

model structures and underlying analytic assumptions. However, in half of

the studies the sensitivity analysis was limited to deterministic methods. In

addition, most of the studies employed rather simple approaches to

sensitivity analysis, with 22/30 using one-way types rather than more

sophisticated multi-way or scenario analyses.

With regard to type of parameters and related uncertainties, effect size (e.g.

number of people becoming active) and direct costs of intervention (e.g.

technical staff) were the input variables most frequently used (21/30) for

sensitivity analysis. Goyder et al. (2014) applied two alternative modelling

approaches, evaluating a large number of scenarios. In particular, results

from a short-term model comparing the quality of life of participants and their

use of health care resources at two different time points (at three and nine

months), were compared to those from a long-term epidemiological model.

The latter more complex modelling approach was developed so that

differences in PA and energy expenditure were mapped onto effects on

mortality reported in the epidemiological literature. This approach not only

served to test the results of the short term model. It also allowed the

mediating effect of PA on health to be formally represented and showed the

dependence of modelling results on strong assumptions made about PA-health dose-response relationships.

A number of modelling assumptions were tested in seven studies (Cobiac et al., 2009, Frew et al., 2014, Gusi et al., 2008, Munro et al., 2004, Nshimyumukiza et al., 2013, Roux et al., 2015, Stevens et al., 1998) which included those related to: decay of effect size over time (e.g. number of new participants), implementation, adherence, recruitment, participation and compliance rates. In four studies (Beale et al., 2012, Moodie et al., 2009, Munro et al., 2004, Sevick et al., 2000), costing assumptions, such as attribution of overhead costs, were explicitly considered within sensitivity analysis, while in only study (Dallat et al., 2014), time lags to disease were included. Finally, results were also tested for sensitivity to variations in discount rates and time horizons, in only three (Anokye et al., 2014, Dallat et al., 2014, Gulliford et al., 2014) and two studies (Cavill, 2011, Frew et al., 2014), respectively.

2.4.4 Critique of the analytical methods used for economic evaluation

2.4.4.1 Estimation of effectiveness

Different research approaches exist to evaluate the effectiveness of interventions (McGovern et al., 2001). The randomised controlled trial (RCT) is regarded as the gold standard, as it is the most scientifically robust method of hypothesis testing (Last, 2001). In this type of study, participants are randomly assigned to one of the identified intervention options. The random

allocation allows for balancing baseline systematic differences between participants that may affect the outcome. In the context of healthy behaviours, for example, this is important because participants more ready and willing to improve behaviour may self-select themselves into the intervention, so inducing bias in the analysis. In practice, however, a RCT cannot always been conducted due to financial, practical and ethical constraints (Craig et al., 2008). Quasi-experimental methods have been recommended in cases where randomisation is not possible. Choice of method is dependent of the type of data available and context. In public health settings, the use of natural experiment frameworks have been suggested (Craig et al., 2012).Through this method, "nature", i.e. factors outside the control of investigators, determine whether individuals are under the control of intervention conditions. Other non-experimental methods, which can be used in combination to strengthen causal inference in natural experiment studies, include interrupted time-series. Through the use of multiple pre/post exposure measures, this method allows to control for secular changes. Lower levels of study designs include simple pre-post assessments where only two, unadjusted measures are used for evaluation.

Baseline systematic differences are not the only source of bias in estimating effectiveness. Loss to follow-up, referred to also as longitudinal selection, represents a main issue, as it can severely compromise the internal validity of results. This represents patients who at one point after baseline have become lost for a certain reason and on whom no information/data is longer available. The severity of this effect is dependent on the reasons for dropping

out, as well as the proportions of participants. Mechanisms of missingness have been classified in the literature (Briggs et al., 2003, Faria et al., 2014) as data missing completely at random (MCAR), covariate-dependent MCAR, missing at random (MAR) and missing not at random (MNAR). A number of post hoc methods is available to address the issue of longitudinal selection and their appropriateness depend on how plausible the assumptions regarding the missing information are.

Differences in intervention effects can also be explained by participant characteristics who react to exposures in different ways (i.e. heterogeneity of effect). This is particularly important when evaluating population-level programmes, as more than one group is exposed to the intervention. Furthermore, changes in one domain of behaviours (e.g. start attending the gym) may affect other domains (e.g. increase in leisure time exercise during the week increases sedentary time on weekends) or other related behaviours (e.g. increase in food intake). Therefore, it is important to assess whether synergistic or compensatory effects take place due to the intervention and account for them when estimating the effect of interventions.

2.4.4.1.1   Effectiveness analysis design

Fourteen of the 21 economic analyses of universal programmes were based on direct estimations of effectiveness. The remaining seven sourced effectiveness input parameters from evidence syntheses (i.e. meta-analyses) or literature reviews. Six studies were conducted alongside trials, namely, three cluster RCTs (De Smedt et al., 2012, Golsteijn et al., 2014, McEachan

et al., 2011), one RCT (Haas, 2006) and one prospective controlled study (Chen et al., 2008) and one single-arm trial (Vestergaard et al., 2006). One study (Guo and Gandavarapu, 2010) analysed national survey data using a transport econometric model (i.e. spatial seemingly unrelated regression). Sutherland et al. (2016) conducted effectiveness analysis on accelerometer data.

The remaining five studies were based on analysis of pragmatic programmes implemented at population-level, where no prior research design and formal evaluation were possible or provided (Cavill, 2011, Frew et al., 2014, Montes et al., 2012, Peterson et al., 2008, Wang, 2005). These studies adopted ad hoc data collection processes, pooling samples from the target population at two time points during their implementation.

Cavill (2011) assessed an intervention aimed to stimulated increases in levels of cycling in six towns across England. Automatic and manual cycle counters were positioned in identified locations to monitor the number of cyclists. Telephone surveys used quota samples to measure whether a change in cycling levels occurred during the four year of programme implementation. In a similar fashion, Montes et al. (2012) analysed a mass recreational programme rolled out in South America and United States through which street were temporarily closed to motorised transport.

Wang (2005), who assessed the return on investment of building bike and pedestrian trails in Lincoln (USA), also based their analysis on census

reports carried out by volunteers who counted cyclists, walkers and outdoor exercisers. Peterson et al. (2008) assessed the cost-effectiveness of a mass-media campaign in school year adolescents in Delaware (USA), while Frew et al. (2014) analysed a universal offer of off-peak access to public leisure centres. Both these analyses were carried out after the programme was rolled out and based on surveys conducted on convenience samples. All these studies used a before/after approach to estimate change in PA behaviours (i.e. parallel trend assumption).

2.4.4.1.2   Longitudinal selection bias

Only four analyses stated to have taken any action in regard to issue of selection effects. Golsteijn et al. (2014) employed linear interpolation methods for outcome data missing in the second of the three data points, and last observation carried forward for the second and third measurements. Vestergaard et al. (2006) stated to analyse only cases with complete observations, while Guo and Gandavarapu (2010) and (Groessl et al., 2016) simply stated to have accounted for missing data.

All the remaining studies did not report on potential issues occurred, with baseline self-selection effects being discussed only in a minority of cases. Worthy of note, Frew et al. (2014) based their assessment on a complete case analysis approach. Participant data were collected in 19 of the 52 council-run leisure centres over an 8 week period. Of the 2556 participants providing baseline data, 797 provided follow-up survey data. It is on this sub-sample that effectiveness analyses were performed.

2.4.4.1.3 Heterogeneity of effect

Except for Haas (2006) that did not detail on this aspect, all controlled trials accounted for heterogeneity of effects, including adjustments for socio-demographic variables, such as age, gender and measures of socio-economic status. Among the other types of studies, Vestergaard et al. (2006) also accounted for participant-level heterogeneity, while the remaining papers did not include details on this aspect.

2.4.4.1.4 Spill-over effects

Only a handful of studies addressed or documented spill-over effects within their analyses. Cobiac et al. (2009), who evaluated the impact of a mix of PA interventions, mentioned synergistic effects possibly occurring as a result of implementing them at the same time. Gulliford et al. (2014), who performed an EE of a universal strategy to promote PA in primary care, acknowledged of having assumed no social multiplier effect, whereby the impact on one person taking on more PA might influence others' around them.

The occurrence of such an effect was instead argued within the discussion of findings by Moodie et al. (2009), who evaluated an active transport programme for primary school children, maintaining that their estimates were conservative as not including the impact of the programme on the wider student population, parents and wider community.

Barrett et al. (2015) assumed no compensatory effects by children during other times of the same day documenting with available evidence. These

authors also argued that the physical education policy subject of EE could be able to change the social norm about being active at school, while also increasing teachers' ability to promote PA. Cavill (2011) used survey data to document whether any increase in PA by participants of the cycling-based programme events was offset by a corresponding decline in other forms of activity (e.g. increased sedentary time).

The only study formally accounting for compensatory and synergistic effects through quantitative analyses was that by Guo and Gandavarapu (2010). This study evaluated the impact of a change in the built environment, namely, adding sidewalks to all roads, applying an econometric model first proposed by Anselin (1988). This model consists of a system of two linear regression equations, where daily vehicle miles travelled and miles walked or cycled by individuals are jointly modelled as a function of changes in defined independent built-environment variables (i.e. neighbourhood, regional accessibility and weather measures), while controlling for a number of additional variables.

2.4.4.2 Modelling of impact

Twenty-one studies employed modelling techniques to extrapolate intervention effects over time (i.e. beyond last assessment) in 22 analyses. Ten of these analyses used untimed modelling frameworks. Aggregate-level modelling was applied in nineteen studies. Three studies used freely available off-the-shelf tools, eight were based on comparative risk assessments and two on individual-level decision-analytic models (Goyder et al., 2014, Nshimyumukiza et al., 2013). Among aggregate level approaches,

Markov modelling was used in the majority of cases (17/19), with only one decision tree Pringle et al. (2010) and one multiple cohort lifetable approach Cobiac et al. (2009). The two individual-level models were also structured as state-transition Markov chains (MCs).

2.4.4.2.1 Heterogeneity of impact

The majority of models were designed to propagate intervention-generated changes to "average" groups of individuals, and around half of them accounted for socio-demographic characteristics including age and gender. Six analyses were based on models able to capture heterogeneous effects according to baseline levels of PA. In particular, only the models by Cobiac et al. (2009) and by Dallat et al. (2014) considered population-level distribution of PA categories, but did not report the model structure. Frew et al. (2014) and Over et al. (2012) modelled three levels, while the two papers (Roux et al., 2015, Roux et al., 2008) presented a Markov model with four PA states matching the current, at the time, classification suggested by PA recommendations in the United States.

According to the presented diagrams and model descriptions, none of these models allowed for full transition between PA states. This limited the possibility of representing fluctuations in PA habits, which can be relatively unstable over short periods of time or during sensitive life phases (Van Dyck et al., 2017).

Further, these models were designed such that intervention-generated improvements in PA could *only* be translated as increases in PA category. More specifically, members within the identified non-active categories of PA could only move upwards or stay in their baseline categories, and the impact of the intervention on already active individuals was assumed neutral (i.e. they would remain active).

Changes in PA categories were defined as a function of increases in intensity and frequency of PA, using METs which represent measures of energy expenditure. Relative risks (RR) parameters for PA categories were sourced from literature reviews, with linear interpolation methods being used when no relevant estimates were found. Four papers disclosed details regarding time lags of beneficial effects of PA. Anokye et al. (2014) and Barrett et al. (2015) assumed a one and two year periods, respectively, for health benefits to start accumulating. Cavill (2011) assumed that individuals would benefits gradually over time, taking five years to reach maximum level. Dallat et al. (2014), who used an existing tool (i.e. PREVENT), applied a range of time lags (from one to fifteen years) to accommodate that fact that benefits may not emerge instantaneously. These estimates were sourced from epidemiological studies on different age groups.

2.4.4.2.2 Behaviour change maintenance

Nine studies of the 21 studies extrapolating intervention effects over time (e.g. after follow-up) did not disclose any detail regarding measures or assumptions on maintenance of behaviour change. All the remaining studies

assumed that the intervention effect would remain constant over time, at different rates.

For their base-case analyses, five models assumed that intervention effects would not decay over time, one study assumed that a residual 80% would remain in place (Munro et al., 2004), two studies a 50% (Pringle et al., 2010) , and Over et al. (2012) used an estimation method (Jacobs-van der Bruggen et al., 2007) which produced an estimated 25% of residual effect. Goyder et al. (2014) imposed an assumption of immediate rebound (i.e. 100% decline of intervention effect) after two years from the simulation started, while (Roux et al., 2008) modelled a constant decline in effect (from 33% to 50%) after that time period. None of the reviewed models accounted for heterogeneity of effect decay according to baseline characteristics.

2.4.4.2.3   Range of diseases

Thirteen studies modelled the impact on chronic diseases and conditions. Except for Nshimyumukiza et al. (2013), who focused on Osteoporosis and fracture events, the remaining twelve models selected different combinations of diseases, with type II diabetes being included in all of them. As well as type II diabetes, up to two more or less broad cardiovascular conditions (ischaemic stroke, stroke, myocardial infarction, coronary heart disease, cardiovascular disease), up to two types of cancers (breast, colorectal and colon cancer) were selected by modellers.

In terms of mental health outcomes, depression was also chosen by Amarasinghe (2010) and Gulliford et al. (2014) who modelled this condition as a comorbidity with other four diseases. None of the studies modelling chronic diseases accounted formally for Obesity outcomes, to avoid double-counting of benefits. Analyses based on direct measurement of health outcomes included Falls (Haas, 2006), and mobility disability (Groessl et al., 2016), functional ability (Vestergaard et al., 2006), depression (Isaacs et al., 2007) or obesity-mediated reductions in health-related quality of life (Barrett et al., 2015, Guo and Gandavarapu, 2010, Sutherland et al., 2016).

2.4.4.2.4 Non-health effects

In the majority of reviewed studies, benefits were measured in QALY or DALY terms and generated as functions of reductions in disease risks. Consideration of broader measures of outcome depended on the perspective taken. In the paper by Guo and Gandavarapu (2010) air quality benefits derived from changes to the built environment were accounted for. Beale et al. (2012) and Cavill (2011), who also modelled environmental interventions, included comfort and security, traffic congestion and productivity outcomes. The latter outcomes were also considered in the analysis by Golsteijn et al. (2014), (Roux et al., 2008) and Roux et al. (2015), who assessed the cost-effectiveness of interventions in an American context, and included gains in QALY calculated from increases in PA level. Frew et al. (2014) estimated the value of the programme to participants using a contingent valuation methodology.

2.4.4.3   Economic analysis

A broadening of perspective from a narrow health care sector viewpoint to include economic effects on other areas of the public and private sectors, as well as in terms of health inequality impacts, have been recommended as good practice (Weatherly, 2009). Relevance of these aspects is dependent of the type of intervention, and especially in regard to the latter, on whether universal or non-targeted approaches to health promotion have been adopted.

Across the 21 articles, universal strategies were the subject of 32 economic analysis. Eleven studies applied a combination of frameworks. CUA was the most used framework, followed by CEA (9), CCA (8) and CBA (5). Ten studies considered time horizons equal or longer than 10 years, with lifetime being the most used, while the remaining analyses selected time horizons equal or shorter than 5 years.

All the EEs used current practice or no-intervention scenarios as comparators, and the majority of studies followed national guidelines to determine what discount rate to apply. Five of the 21 studies did not disclose what perspective was taken for EE. Eight studies stated a health care or health system perspective, two the remaining stated "wider", public payer, public health or societal perspectives.

2.4.4.3.1   Inter-sectoral costs and consequences

Non-health costs

Four interventions concerned modifications to the physical environment. Cavill (2011) did not detail or describe the costing of the cycling intervention subject of their EE, merely mentioning use of the allocated budget in the first three years of project delivery.

Except for Guo and Gandavarapu (2010), who considered only the building of infrastructures (sidewalks), the studies included both construction and maintenance costs. The study authors assumed an even allocation of these costs throughout the supposed building life cycle, with only two studies (Montes et al., 2012, Wang, 2005) including out-of-pocket costs to potential users. In particular, both these two studies, which were implemented in the American continent, accounted for equipment costs (e.g. sport shoes). In doing so, they broadened the perspective from that of the authority "delivering / administering" the intervention to a wider (i.e. reported as a public health) perspective.

Of the eight studies evaluating PA promotion interventions based on marketing strategies, six studies (Cavill, 2011, Nshimyumukiza et al., 2013, Peterson et al., 2008, Pringle et al., 2010, Roux et al., 2015, Roux et al., 2008) did not disclose how the interventions were costed nor provided any relevant detail. However, according to what was reported, Pringle et al. (2010) included average costs of implementing the interventions, while Cobiac et al. (2009) assumed that the interventions were operating under "steady-state" conditions.

Roux et al. (2008) and Roux et al. (2015), who evaluated and modelled a combination of different interventions, reportedly included all the costs associated with the interventions, which were determined through direct communications with the authors of the original investigations and a review of manuscript protocols. Peterson et al. (2008) did not describe the costs but mentioned to have accounted for both production and placement costs related to their mass-media campaign. Golsteijn et al. (2014), who based their strategies on written and Web-based information (i.e. leaflets), included promotion and development costs, namely costs for invitations and gathering of environmental information, respectively.

Travel costs were included by Cobiac et al. (2009) and Golsteijn et al. (2014), with the former also accounting for the additional time spent by participants for the increased PA level and the latter estimating costs for productivity loss. Time spent by participants was also included and valued by Roux et al. (2008) and Roux et al. (2015), together with expenses for sport equipment. Twelve studies evaluated interventions based on communication strategies, that is, advice, behavioural training or counselling. All of the studies included the cost for consultation by the professional (i.e. typically the time spent). Except for McEachan et al. (2011), who implemented a strategy based on PA facilitators on the workplace, all of the eight studies evaluating face-to-face interventions based outside a primary care or health care setting (e.g. community-based) included the costs related to the hosting facilities.

As for costs related to the development of the intervention, only Sevick et al. (2000) included and detailed the resources used for developing the intervention (i.e. preparation of class contents). Besides the interventions assessed by Cobiac et al. (2009), (Roux et al., 2008) and (Roux et al., 2015), only McEachan et al. (2011) included the perspective by participants, by accounting for their opportunity cost in terms of (valued) time and travel resources employed to attend the sessions.

Of the 11 studies evaluating interventions providing access to exercise or sport opportunities (typically exercise classes) delivered within dedicated facilities (e.g. gyms, leisure centres), four did not include venue costs. In particular, Babey et al. (2014), Barrett et al. (2015), Sutherland et al. (2016), and Chen et al. (2008), who evaluated school-based strategies and a hospital-based programme, respectively, assumed no incremental costs incurred for hosting the intervention.

Six studies accounted for costs borne by participants and their families, voluntary staff and the hosting authority. Babey et al. (2014), Chen et al. (2008), Isaacs et al. (2007) and Sutherland et al. (2016) accounted for costs for enrolling in the programme, such as equipment, childcare, travel and the time spent. Time spent was also considered for the calculation of opportunity cost by volunteers in the study by Moodie et al. (2009) who evaluated a walking bus programme for elementary pupils.

Frew et al. (2014), who took a health care perspective to assess cost-effectiveness, applied the same approach of the other pragmatic programmes (Cavill, 2011, Montes et al., 2012, Peterson et al., 2008, Wang, 2005), that is, a budget expenditure approach. Frew et al. (2014), however, applied a changing annual usage rate (i.e. proportion of participants attending the leisure centres, n~100,000) of 50-100% to account for variations in per-participant cost over time, using a triangular distribution.

Non-health consequences

Only seven studies addressed or estimated the economic impact of the intervention on public sectors other than health. Of these, five studies (Beale et al., 2012, Cavill, 2011, Guo and Gandavarapu, 2010, Montes et al., 2012, Moodie et al., 2009) promoting changes in modes of travel (i.e. walking and cycling) accounted for effects on the transport sector, such as comfort and security, user amenity, traffic decongestion and accidents. Impacts on the environment sector were discussed and estimated in the studies by Montes et al. (2012) and Guo and Gandavarapu (2010), respectively, who included reductions in air and noise pollution outcomes through decrease in the use of passive modes of transport or commuting (e.g. cars).

Changes in productivity (private sector) were assessed in the EEs by Beale et al. (2012), Cavill (2011) and McEachan et al. (2011) who accounted for variations to absence by workers due to ill health. This outcome was of particular relevance in latter paper where a work-place based interventions was evaluated from an employer perspective.

71

Other more complex constructs, such as recreation, social capital development and contact, which could be put under the broad umbrella of social well-being, were discussed in the papers by Montes et al. (2012) and Munro et al. (2004). In particular, the latter authors acknowledged the impact that an intervention involving older adults to take part to exercise classes could have on both a widened network of social contacts and the voluntary contribution that this age group might make in terms of caring (e.g. childcare and caring for relatives).

2.4.4.3.2 Equity considerations

None of the 21 studies incorporated equity considerations into EE formally (i.e. using quantitative methods). According to what was stated by their authors, the main reason for not conducting sub-group analyses was limitations in effectiveness data and sample size.

Nine studies addressed equity considerations qualitatively, either via discussion of findings, or in more structured fashions. The three school-based studies assessing the impact of the intervention on general populations of pupils (Babey et al., 2014, Barrett et al., 2015, Moodie et al., 2009) used a second-stage filter analysis framework to address equity implications related to implementation or scaling up of the intervention. In particular, they discussed barriers represented by differences in the availability of spaces (Babey et al., 2014), or implementation modalities (Barrett et al., 2015, Moodie et al., 2009) between schools and how these could exacerbate existing disparities.

Comparably, Beale et al. (2012) and Vestergaard et al. (2006) made explicit concerns about whom could benefit from the intervention. Vestergaard et al. (2006) highlighted the importance of transport barriers to the intervention for potential participants, while Beale et al. (2012) proposed ways to mitigate this issue, in the context of environmental interventions, by advocating for alternative designs and modes of programme delivery or even subsiding or incentivising access. Golsteijn et al. (2014), Goyder et al. (2014), (Roux et al., 2008) and Frew et al. (2014) acknowledged the limitations of assuming an average effect change for the whole sample, while justifying the lack of relevant sub-group analyses with insufficient statistical powers (Frew et al., 2014, Roux et al., 2008).

2.4.5 Main findings

Overall, PA interventions appeared to be optimal alternatives, compared to "do-nothing" options. Universal strategies presented a large degree of heterogeneity in terms of the analytical methods applied for EE. A number of methodological challenges have characterised these studies. In particular, the review has raised considerations related to the methods used to estimate effectiveness, modelling of impact and economic analysis. Many of these issues, though not exclusive to universal programmes, did characterise the EE of these initiatives, especially when no prior research design was possible.

The review revealed large variability in study design. The majority of studies, especially among those assessing universal strategies, were not based on experimental evidence, but relied on before-after analytical approaches. PA

promotion initiatives were generally found to be cost-effective alternatives, particularly large-scale programmes, which took the form of universal approaches to intervention in the majority of cases.

Reporting of analytical methods was often incomplete and methodological issues were dealt with mainly through unprincipled approaches. Particularly among EEs assessing the impact of universal strategies, the large majority of effectiveness analyses were implicitly or explicitly based on relatively strong assumptions regarding selection effects.

A range of modelling approaches were used in previous studies to extrapolate intervention effects beyond follow-up periods and link changes in PA to health outcomes, using aggregate-level methods in the large majority of cases. The issue of heterogeneity was addressed only in part, with few studies accounting for population-level impacts. While one model (Roux et al. 2008) accounted for natural trends in PA levels over time, no previous model structure could accommodate fluctuations in PA states, therefore, limiting the possibility to model them formally.

The review also revealed that different combinations of chronic diseases and conditions were selected to be modelled in comparable populations. Consideration of consequences, other than health gains due to reductions in diseases risks, including non-health effects (e.g. reductions in traffic congestion), was limited to a minority of models. Furthermore, the issue of decay of effectiveness was addressed only in a small proportion of studies

that mostly assumed constant and homogeneous residual intervention effects to last for the whole time horizon.

In terms of economic analysis, standard approaches to EE were adopted in the majority of applications: a health care sector perspective over a lifetime. Among universal strategies, inter-sectoral costs and consequences were considered mostly in EEs assessing the impact of changes in the built environment (e.g. building of new cycle trails). Finally, the review revealed that the issue of health equity was essentially ignored. Those few analyses which considered the impact of the intervention on health inequities did so only qualitatively, through discussion of the implications of implementing the intervention and proposing alternative intervention delivery options.

2.4.6 Limitation of the review

While the review was deemed comprehensive, it might not have been exhaustive in capturing the methodological challenges characterising the broader field of promotion of healthy behaviours. Having focussed on the EE literature, review of the issues related to the estimation of effectiveness was limited to what was reported in the respective published papers. Therefore, what was found in terms of paucity of appropriate methods of analysis, cannot be generalised beyond this literature, which in turn represents only a limited proportion of the whole evidence base. Furthermore, the review scope was restricted to PA promotion only and a broader review including other healthy behaviours would have allowed to obtain a more accurate picture of the current state.

2.5 Identified methodological issues

The systematic review revealed that a number of methodological shortcomings characterise the current literature of EEs of PA promotion interventions. Considering the relevance of the issues of health equity to public health decision-makers and the apparent gap in this literature, both in term of empirical evidence and methodological development, this was selected as the main outstanding issue to address in this thesis. The review also showed that the issue of behaviour change maintenance and related extrapolation methods used for EE are currently based on often unrealistic assumptions, therefore warranting further methodological research.

Focussing on the empirical evidence for PA promotion, the review showed that key issues especially related to longitudinal selection effects, models' structural assumptions and perspective for EE were not addressed in the reviewed studies. Choice of these methods has the potential to influence the economic decision, and testing them can shed some light on the sensitivity of economic results which is important to adequately characterise the uncertainty surrounding the decision (Husereau et al. 2013). To test these methodological and structural assumptions, as well as demonstrate the decision model's applicability, an illustrative case study was used.

Therefore, as mentioned at the end of the previous chapter, these identified methodological challenges are addressed alongside the EE of the case study programme. The issue of selection bias due to second stage survey non-

response is addressed in the second part of next chapter, following an analysis of effectiveness of the intervention. Chapter 4 focuses on the two modelling shortcomings of population-level impact (which encompasses aspects related to the heterogeneity of impact across population subgroups of mostly healthy individuals and modelling of health inequality) and behaviour change maintenance over time.

Following a base-case CEA, the results of which are presented in Chapter 5, Chapter 6 explores the issue of perspective for EE. Specifically, two alternative decision-making perspectives are investigated, namely: a Health and Wellbeing Board (HWB) and a Local Authority.

In the second part of chapter 6, four modelling and methodological groups of assumptions that underpinned existing economic models, and which are related to the addressed issues, are tested. These assumptions regard: 1) mechanisms of second-stage survey non-response 2) range / combinations of modelled diseases 3) behaviour change maintenance over time 4) perspective for EE.

3   Chapter

Estimation of behaviour change

3.1   Chapter outline

As identified in the previous chapter, the issue of longitudinal selection bias due to survey non-response has been largely unaddressed in previous EEs, particularly within studies evaluating population-level programmes. This represents a main issue because economic evaluation results depend on how the programme is effective at changing PA behaviour and a small difference in estimated effects can have large implications for population-level cost-effectiveness. Moreover, the data collection process cannot be typically designed or conducted as in RCTs, so participants cannot be followed-up on an individual basis. Rather, as highlighted in the review, non-probability sampling approaches are usually employed to collect longitudinal data and this has to be considered when conducting the effectiveness analysis. In order to explore and represent the potential implications of different approaches to address this issue, a case study was used.

The issue of longitudinal selection bias represents the methodological focus of this chapter which also includes an assessment of the distributional effectiveness of the case study programme, to enable an EE in chapter 5.

Section 3.2 describes the case study. Section 3.3 describes the methods used to analyse the case study data, starting with definitions and

identification of the variables. In sub-sections 3.3.4, details on the econometric methods used to estimate the distributional effectiveness of the case study programme are provided, including the identification of two measures of change in PA, which are derived from the case study data. The following sub-section 3.3.5 is dedicated to the methods used to handle selection bias. Section 3.4 provides an overview of the descriptive results, while section 3.5 reports the effectiveness results obtained using a pragmatic approach to second-stage survey non-response. Section 3.6 explores three alternative mechanisms of non-response scenarios. Section 3.7 summarises the main findings and implications.

3.2 The case study

3.2.1 Decision-making context

In 2012, Sport England launched a funding competition named 'Get Healthy, Get into Sport'. Leeds City Council and NHS Leeds/Public Health (at the time, before Public Health England began operating in April 2013) submitted a joint proposal, named 'Leeds Let's Get Active' (hereinafter LLGA). In March 2013, it was confirmed that the proposal had successfully secured funding from Sport England, which was matched by Public Health England. The latter agency also provided additional financial support to extension of the programme until the end of 2016. The local City Council provided in-kind support, mostly in the form of staffing and facilities (i.e. leisure centres). In 2015, after around two years since the programme was rolled-out, the City

Council commissioned the EE of LLGA programme, which provided a case study for this PhD project.

LLGA was a pragmatic programme which was designed based on a programme previously implemented in Birmingham called 'Be Active' and evaluated by Frew et al. (2014). Other than research studies on benefits of PA and potential cost-savings deriving from improvements in PA, the research by Frew et al (2014) was the only piece of evidence which the City Council provided as base and support for investing in LLGA (see Appendix K).

3.2.2 Programme contents

LLGA was a universal programme to promote PA in the general population. LLGA offer consisted of providing free access to off-peak leisure centre-based exercise sessions to all city residents. In order to encourage physically inactive residents, and especially those from low socio-economic backgrounds, LLGA sessions (i.e. the service) were provided in 17 City Council leisure centres located in the most deprived areas of the city. Key intervention ingredients were therefore the removal of financial barrier to gym membership and geographical proximity of the leisure centres. LLGA service included the use of free weight areas, swimming pool and fitness classes. This form of LLGA ran until December 2016.

According to City Council reports, a financial budget of £1,525,000 was allocated for implementation of LLGA over a period of 39 months of

programme duration (from end of September 2013 to end of December 2016). From this point, LLGA changed from offering a free service to small subsidised charge for use, but no data were made available on this period. It is the 39-month period of free service that will be the focus of the present work.

Marketing and reach

A partnership between the City Council and Public Health produced a marketing campaign that was carried out in the six months before the launch of the programme. Using traditional approaches (e.g. leaflets and bus shelters, see appendix C) and digital platforms (e.g. banners, messages and emails to prospective and existing leisure centre customers), the promotional activity aimed to attract target groups. Target groups were physically inactive residents from deprived areas of the city. This activity was based on communication strategies centred on an offer of a supportive and welcoming environment. A website (Active Leeds, 2013) was also created to promote LLGA activities and messages, as well as for data collection purposes.

LLGA sessions were scheduled at certain times of the day, mostly for an hour a day during off-peak hours (i.e. majority of sessions in the morning and early afternoon). According to City Council reports, this was due to a number of reasons. First, the City Council wanted to ensure sufficient capacity to incorporate new facility-users (i.e. new members using the leisure centres) and that programme sessions could fit into routine leisure centres' session

timetables. Secondly, because off-peak sessions were thought to be most appealing to target groups (e.g. unemployed residents).

Finally, sessions were selected to correspond to those times that were likely to have the lowest revenue impact, in terms of loss of earned income from existing fee paying customers. To mitigate this risk, programme staff put efforts on offering additional paid sessions to new members. In fact, programme participants also had the opportunity to become leisure centre customers and attend the other routine sessions available in the facilities outside LLGA, at the standard price. As with new members, an induction course was offered to all participants.

3.2.3 Data collection process

Collection and gathering of data were carried out by programme staff. No eligibility criteria were defined for a city resident to register to LLGA and access the service. Residents who were already leisure centre members could also sign up, and consequently, become a LLGA participant. Registration could be done either in person at the leisure centre receptions, or on-line through the programme website.

On providing their personal details, participants were individually assigned an electronic card. There was no restriction imposed to participants in terms frequency or regularity of access (e.g. participants could attend sessions at any time during the programme), or number of programme sessions they could attend. Service use was electronically monitored by means of the LLGA cards swiped at the leisure centre gates. Participants were also

surveyed through self-report questionnaires twice. First, at the time of their individual registration to the programme (i.e. baseline). Later during the programme, participants were asked to complete the survey questionnaire a second time. Second-stage survey data were collected during a series of "follow-up weeks", which were implemented roughly every six months during the first 30 months of the programme. As a consequence, both the period of granted access to the service (i.e. from baseline to end of 2016) and survey assessment (i.e. from baseline to second survey time point) were not uniform but varied by participant.

During implementation, because of a shift of interest by decision-makers to include additional outcomes potentially associated with the programme, the survey questionnaire was replaced by one that included additional socio-demographics (hereinafter "lifestyle questionnaire", see appendix D.2). This resulted in only part of the measures matching between those who registered to the programme before (i.e. cohort 1) and after 31.03.2015 (i.e. cohort 2). The modality of data collection also changed from being in person, in majority of cases during cohort 1 phase, to mostly on-line in cohort 2 phase.

Issues arose in regard to the "follow-up weeks". In particular, the programme administrators did not keep record of how many participants were surveyed a second time and what participants of those surveyed did not provide a response. As a result, retrospectively, it was not possible to know reasons for missingness in the data.

Outcome measures

Programme staff, with the support of an academic partner (Leeds Beckett University), who conducted a programme evaluation centred on interim reports from qualitative and descriptive research, identified the questionnaires to employ. At the moment of their registration, participants were asked to provide details on age, gender and residential postcode (mandatory fields) through a registration form.

During cohort 1 phase, participants were surveyed using a modified short-form version of the IPAQ questionnaire (see appendix D.1). This questionnaire have been subjected to validity and reliability tests in many countries including the UK, and found to have reasonable measurement properties for national monitoring purposes (Craig et al., 2003). However, as the other self-report measures it is prone to recall bias. Indeed, a more recent systematic review revealed that there is weak evidence to support its use as an indicator of PA, as it overestimates PA as measured by objective criterion by an average 84% (Lee et al. 2011).

This IPAQ questionnaire includes items on the time spent in activities of various intensity (i.e., vigorous, moderate, walking, sedentary time). Within the modified IPAQ questionnaire used to evaluate the impact of LLGA, an active day was defined as a day with at least 30 minutes of at least moderate PA. Unlike the original questionnaire, the first single-item question was focused on the number of active days over the previous seven, which was

used as the primary outcome for analysis. During cohort 2 phase, participants were asked to complete the lifestyle questionnaire which had in common with the modified IPAQ questionnaire only its first question.

Data and information sharing

A Data Processing Agreement was established between the data provider (City Council) and the data processor (University of Leeds, see appendix E). Through this agreement, restrictions were imposed in terms of information provision. In particular, details on participants' residential postcodes were not disclosed by the data provider. Instead, data on the Index of Multiple Deprivation (IMD) status were made available (Department for Communities and Local Government, 2016). LLGA data were shared via secure server. Email communications and a number of in-person meetings with the staff were held, where project objectives, progress and issues were discussed. This was to ensure that the planned evaluation would best meet decision makers' information needs.

3.3   Methods

The analyses conducted here focussed on the adult population, defined as residents aged 16 years and over. Leisure centre rules required individuals aged less than 16 years old to be accompanied on a one-to-one basis by their parents or legal guardians. For comparison with the evaluation by Frew et al. (2014) and previous research studies, individuals aged under 16 years were therefore excluded from the following analyses. No other inclusion or

exclusions were applied. All analyses were conducted using Stata software version 14 (Stata Corp, Texas, US).

3.3.1 Definitions

The case study data set is composed of survey questionnaire and card swipe data.

Participant: a participant is defined as any adult who registered to the programme and provided at least basic socio-demographic data. Basic socio-demographic data include age, gender, IMD status and baseline PA level (i.e. number of active days, NAD, over the seven days prior to registration - questionnaire data). Information and data were only available on programme participants.

Survey respondent / non-respondent. Participants for whom two successive survey outcome measurements were available were classed as survey respondents. Otherwise, if only baseline outcome data were available, survey non-respondents. This groups includes participants who were not surveyed a second time (unknown proportion).

Service user / non-user. Service use means the same as access to LLGA sessions (card swipe data). Participants could either have attended LLGA sessions at least once (i.e. service user), or not at all (i.e. service non-user).

Access period: period of time between participant's individual date of registration and 31st December 2016 (programme end).

Service use period: period of time (either corresponding to the entire access period or discrete sub periods) in which a participant used the service for at least one time.

Facility user / non-user. Facility refers to leisure centre (card swipe data). Other than being service users, participants could access the leisure centres at least once in the 6 months prior (pre-LLGA facility use); and/or access at least one leisure centre session outside LLGA during its implementation (outside-LLGA facility use).

3.3.2 *Identified variables*

Table 3.1 below describes the variables identified within LLGA data set.

The dependent variable (PA) was categorised in the same way the current research linking PA and health benefits (Lee et al., 2012) and UK PA guidelines for adults do (Health, 2011a). That is, four categories; inactive=zero, insufficiently active=1 or 2, moderately active=3 or 4, active 5 or 6 or 7 active days a week. An active day is defined as a day with at least 30 minutes of at least moderate PA (in any domain, occupational or non-occupational). However, across chapters, PA will be also treated as a continuous or an interval variable taking discrete values between zero and seven.

Table 3.1 Identified variables

Variable	Description	Notes
Outcome variable		
Physical activity	4 ordinal categories, defined according to the number of active days (NAD) over previous seven days: Inactive= 0, Insufficiently active= 1-2, Moderately active= 3-4, Active= 5-6-7.	For descriptive purposes, considered also as an interval (0-7), as well as a continuous variable
Survey questionnaires		
Age group	3 ordinal categories (years): 1= younger adults: 16-40, 2= middle-aged: 41-64, 3= older adults:>= 65	Considered also as a continuous variable
Gender	Female or male	Reference category=female
Index of multiple deprivation status	4 ordinal categories, 0=Non-deprived LSOA area, 1= Deprived top 20% IMD score; 2= Deprived top 10% IMD score; 3= Deprived top 3% IMD score	Reference category= Non-deprived
Body mass index status*	3 ordinal categories (score): 0= if 18-25 healthy, 1= if<30 &>=25 overweight, 2= if>30 obese*	Reference category=healthy
Registration date	Date of registration to LLGA programme	From 30.09.13 to 31.12.16
Cohort status	Time period of registration to LLGA, before (cohort 1) or after 31.03.15 (cohort 2)	Different survey questionnaires
Ethnic background*	Binary, White British /Irish, not	Reference category= White British /Irish
Diagnosis of chronic disease status*	Binary, healthy, or diagnosed with any chronic conditions or diseases over last 12 months	Reference category= healthy
Education status*	Binary, 0= higher education (diploma/ BSc/ MSc/ PhD) or 1=not	Reference category= higher education
Employment status*	3 categories, 0= full-time; 1= part-time employed, student or volunteer; 2=unemployed, retired or unable to work	Reference category= full-time
Relationship status*	Binary, 0=living alone or 1= not	Reference category= alone
Card swipes		
Service use status	Binary, 0="no LLGA session attendance", 1=at least 1 LLGA session attended"; total number of LLGA sessions accessed	Reference category=service non-user; considered also as a count variable
Weekly rate of service use	Weekly rate of access to LLGA sessions from registration to programme end / or discrete service use periods	Ratio variable – count of sessions divided by n. of weeks
Facility use pre LLGA status	Binary, 0="no pre-LLGA session attendance", 1=at least 1 other session attended prior to LLGA"	Reference category=pre-LLGA facility non-user
Facility use outside LLGA status	Binary, 0="no other session attendance", 1=at least 1 other session attended outside LLGA"	Reference category=outside-LLGA facility non-user

*available only for cohort 2 participants, NAD=number of active days; World Health Organization classification of obesity in adults, IMD=Index of Multiple Deprivation.

The majority of survey variables had empty cells. While for certain variables, such as gender, a blank cell could only be interpreted as a missing value, for other variables a missing observation could also represent a zero value. In particular, from the way the question related to ill health status was asked, a

zero value could mean either that the information was not available or that no diagnosis of disease occurred. This variable was kept in only for descriptive purposes, assuming that a blank cell corresponded to a no diagnosis state.

Service use data were available in weekly counts. Average weekly rate of service use was calculated dividing the total number of LLGA sessions accessed during their access period, by the number of weeks between these two time points. Height and weight measures were converted to BMI scores, according to the standard formula weight (kg)/height (m)^2. The WHO classification for adults (World Health Organization, 2006) was used for defining healthy (BMI score 18-25), overweight (BMI score 25-30) and obese (BMI score >30).

3.3.3 *Descriptive analysis*

Summary statistics were calculated to describe the distribution of participants according to baseline characteristics. Service use and weekly rates of access to LLGA sessions were also illustrated. Time-related heterogeneity of service use was represented by discretizing the access period in sub periods and including only service users with access periods of equal length or longer.

Probability of service use was described using RRs. Extent of missing data was described, along with an appraisal of the reasons for missingness. This was within constraints related to the data collection process. Namely, no involvement in the data collection and management processes and not being

allowed to collect additional information on missing values (in accordance to the Data processing agreement).

Statistical significance was set at a 0.05 threshold. For continuous variables, means and standard deviations (SDs) were presented. For categorical variables, relative proportions (%) were displayed. Differences in personal characteristics were tested between sub-samples using analysis of variance (ANOVA) or independent sample t-tests for continuous variables, as appropriate, and χ^2 tests for categorical variables. An informal analysis of residuals followed for significant estimates of categorical variables with more than two levels.

3.3.4 *Distributional effectiveness*

3.3.4.1 Measures of behaviour change

Programme effectiveness was defined as the ability of LLGA to affect change in overall PA. As being involved only retrospectively, the choice of measure of behaviour change was constrained to what the programme administrators had identified. Measures more reliable than self-reported and card swipes exist, but they were not used in LLGA. Objective measures of PA, such as pedometers and accelerometers (e.g. ActiGraph, O'Neil et al. 2014), have been increasingly used to assess the impact of PA interventions, although they remain underutilised to date (Silfee et al. 2018). While the reliability and validity of these measures may be superior to the other types in controlled conditions, accuracy of the PA measurements relies on the location of the

device (e.g. wrist or thigh) and a stable position on the body which may not always be kept in real life settings (Yang et al., 2018).

On the other hand, although it represents an objective measure, using the card swipe as a measure of PA behaviour implies a rather strong assumption: that the participant uses the card to actively attend a gym session at an intensity that increases their heart rate significantly and for at least 30 minutes (i.e. an active day). This remained an assumption that could not be supported by evidence, as further data on participants could not be collected (i.e. due to the restrictions imposed by the Data Processing Agreement).

From a retrospective standpoint, the choice of measure of behaviour change was constrained to what the programme administrators had identified. Given the importance of valid measurement of PA behaviour to be able to assess any effect of LLGA programme, this case study was used only to explore the implications of the identified methodological assumptions (chapter 2) and apply the developed decision-analytic model (see chapter 5).

Change in the frequency distribution of the four PA categories was assessed using two measures of PA behaviour change. The first outcome measure (hereinafter "survey measure") was based on the survey data only, as the change in self-reported PA category (i.e. from the single-item question) observed between baseline and second survey assessment (after registration, "survey follow-up weeks").

The second outcome measure (hereinafter "card swipe measure") was based on baseline survey data on NAD and card swipe data (weekly rate of service use). This measure was calculated as the probability of participants to improve PA category, due to a sustained rate of access to LLGA sessions of at least one time a week. In order for a participant to improve PA category, they needed not to have self-reported themselves as already active (i.e. 5 or more active days a week).

A sustained rate of service access of one time a week, therefore, was sufficient for those participants who at baseline self-reported themselves as being inactive (zero active days), 2 active days (classed at insufficiently active) and 4 active days (moderately active) to move to the respective next higher PA category. For those participants who reported baseline 1 or 3 active days, or for the other non-active to improve PA category by two levels (e.g. from inactive to moderately active), a rate of service access of at least 2 times a week was needed. A constant rate of service access was assumed over the considered analysis periods.

3.3.4.2   Analysis approach

In order to minimise the risk of overestimation of the intervention effect, a conservative approach was taken. In particular, a last-observation carried forward method was applied, such that change in PA by survey non-respondents and by service non-users was assumed zero and included in the effectiveness analysis.

Given the aim of LLGA to encourage especially physically inactive and residents from deprived city areas to become active, and a modelling structure that needs to incorporate the issue of health inequality, a distributional approach to estimation of behaviour change was taken. In particular, an ordered logistic regression approach was applied (i.e. four PA categories, as specified in section 3.2.2), with neighbourhood-level deprivation (i.e. IMD) status and baseline PA being identified as the equity-relevant characteristics. These two variables were interacted, with the resulting interaction term representing the main explanatory variable.

The assumption of proportional odds was first tested using a likelihood ratio-based test of proportionality of odds across response categories, for the overall model. Wald test was used to assess statistical differences between deprivation subgroups for each of the four PA categories. If the model tested positive, an auto-fit procedure was first applied to determine whether the outcome variable met the proportionality assumption. This procedure also guided choice of a model specification which best fitted the data. The response variable was regressed on the identified explanatory variables, within the group of participants. The variables age and gender were kept in the models regardless of their estimated statistical significance or effect size.

3.3.5 *Handling selection bias*

As mentioned above, for only a small proportion of participants a second survey measurement was available. In addition, no record of the number and characteristics of LLGA participants who were surveyed a second time (after

registration) was kept by programme staff. Therefore, it was not possible to discern whether survey data had not been collected on a participant in the first place or were missing due to intentional non-response or data mishandling.

In order to address the missingness issue, alternative scenarios were explored, within the constraints mentioned above (see section 3.2.3). In terms of information available for this analysis, the data provider acknowledged problems in obtaining and sharing complete information on LLGA participants, due to reluctance by participants and lack of resources to pursue follow-up measurements. Reluctance to provide information by participants was of particular concern, as self-selection could have occurred and bias been introduced.

Three formal selection mechanisms scenarios were explored and their results compared with that of the pragmatic approach used for estimation of effectiveness (i.e. last observation carried forward). Selection bias was addressed under assumptions of second-stage survey data MCAR, MAR and MNAR.

With MCAR, average change in PA within survey non-respondents was assumed equal to that within survey respondents (i.e. complete case approach). In other words, the change in PA observed in the data was assumed to be representative of the entire cohort of participants.

Relying on a MAR assumption, a selection on observables approach (Moffit, 1999) was chosen to address this scenario. An auxiliary variable influencing both the selection process and affecting the investigated outcome was identified. Using probit model estimates (see appendix H), an inverse probability weighting method was applied to adjust for selection bias (Seaman and White, 2013). Put simply, all participants were assumed to be surveyed and each observation (case) was given a weight equal to the inverse of the individual probability of being selected, with all weights adding up to one.

A MNAR scenario was analysed using a selection on unobservables approach (Moffit, 1999). In this setting, as in MAR, it was assumed that all participants were surveyed a second time. Using this method meant that survey non-response was assumed to be driven only by unobserved factors correlated to determinants of PA, after conditioning on the identified explanatory variables. Following the approach described by (Wooldridge, 2011), a Heckit two-step estimation was conducted using the least square method. This method allows to correct for the unobserved selectivity (Jones, 2007). The variable cohort, which was believed to have influenced the longitudinal selection process, was chosen as the instrument. This was also used to reduce the risk of collinearity due to the inclusion of age, gender and IMD status in both the selection and structural equations.

3.4 Descriptive results

3.4.1 Reach

Of the residents exposed to LLGA offer (i.e. at least awareness of the programme), 79,115 adults signed up to the programme. For 65.6% of these (n=51,874), basic socio-demographic data were available. This group of 51,874 adults were defined as participants.

Table 3.2 below summarises the characteristics of the sample in terms of baseline socio-demographics, with an indication of the proportion of missing data. According to the observed values, that is, ignoring missing data, the mean age of participants was 38.5 years, the majority were White British, at a healthy weight or overweight, female, not living alone, at least part-time employed and without a higher education degree.

The programme was able to attract a large number of residents, and targeted individuals to a certain degree. Participants were for almost two-thirds insufficiently or completely inactive, a fifth lived in deprived city areas and only seven percent were already attending the facilities prior to implementation of LLGA programme. A separate discussion on missing baseline data is presented in section 3.3.4.

Table 3.2 Participant baseline characteristics

Characteristic	Reference category	N=51,874
	16-40 y	61.5%
Age group	41-64 y	31.5%
	>64 y	7.0%
Gender	Female	62.4%
Index of Multiple Deprivation status	Non-deprived	80.5%
	Top 20%	11.0%
	Top 10%	7.2%

			% m. obs	% m. obs Cohort 2
	Top 3%	1.3%		
Physical activity category	Inactive	29.0%		
	Insufficiently active	37.0%		
	Moderately active	21.4%		
	Active	12.6%	**% m. obs**	**% m. obs Cohort 2**
Cohort status	1, signed up before	62.5%	0%	0%
Body Mass Index status*	Healthy weight	44.3%	87.6%	66.8%
	Overweight	31.6%		
	Obese	24.1%		
Ethnicity*	White British	77.8%	74.7%	32.7%
Education status*	Higher level	39.3%	74.7%	32.7%
Ill health status*	No diagnosis	86.3%	62.5%	0%
Employment status*	Full-time	48.0%	74.7%	32.7%
	Part-time	27.7%		
	Unemployed	24.3%		
Relationship status*	Living alone	41.4%	74.7%	32.7%
Pre-LLGA facility use	No leisure centre access	92.7%	0%	0%

*available only for cohort 2 participants (n=19,438); ill health= diagnosed with a chronic condition over the past 12 months; m.=missing; obs=observations, y=years.

3.4.2 Service use

A total of 191,605 accesses to LLGA sessions were totalised by 23,481 service users over the 39 months of programme duration.

3.4.2.1 Service use patterns

Around half of LLGA participants did not access any LLGA sessions (54.7%, n=28,393), 12.6% attended the sessions only once, and around 5% of participants were distributed between 23 and 780 accesses. Service users accessed LLGA session at a mean value of 5.16 (20.1).

Figure 3.1 below shows the drop-off patterns of service use, defined as the distribution of service users by time from registration to last access to LLGA sessions. Around half of those attending the sessions for at least one time stopped doing so within 6 months from signing up to LLGA. From a 7.6% between 6 and 9 months, the proportion of service users progressively

decreased up to a 0.4% who registered at the beginning of the programme, and waited the last weeks for accessing its free sessions.

Figure 3.1 Service use drop-off pattern
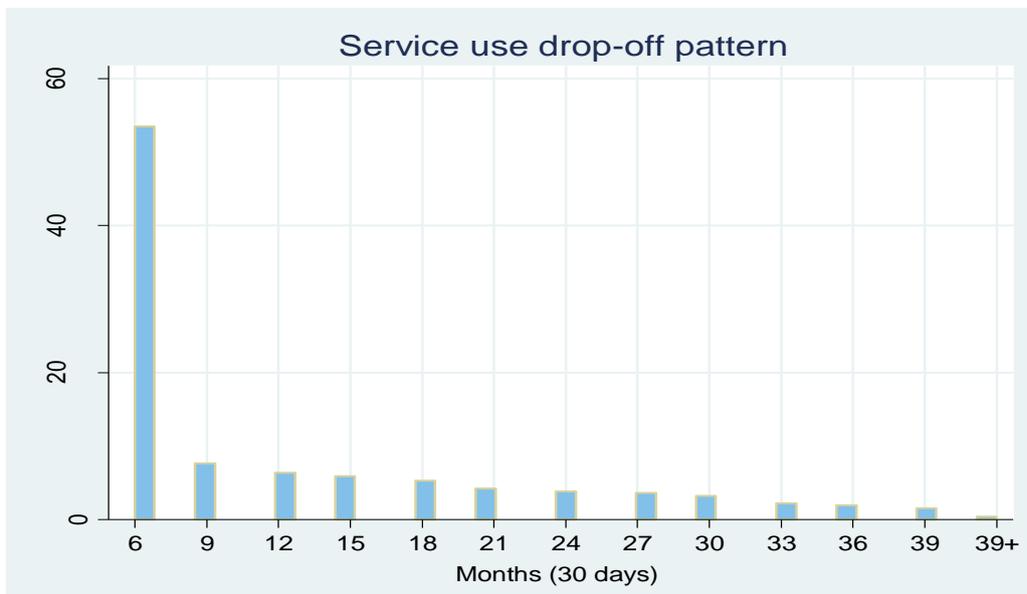
3.4.2.2   Probability of service use

As shown in Table F.1 (see appendix F), the unadjusted probability of using the service at least once was greater for older, males, non-inactive, relative to the respective comparators. After adjustment, the risk was 14.8% and, 5.3% higher for older adults and males, respectively. Relative to the group of physically inactive, being insufficiently active (i.e. 1-2 active days a week) was associated with a 11.8%, while moderately active or meeting the PA recommendations with around a 20% higher probability of service use status.

Weekly rate of service use

From the card swipe data, it was possible to estimate to what extent a habit of attending the free off-peak sessions was established among service users.

To this purpose, Figure 3.2 below shows the distribution of service users by average weekly rate of service use.

Figure 3.2 Distribution of service users by average weekly rate of access
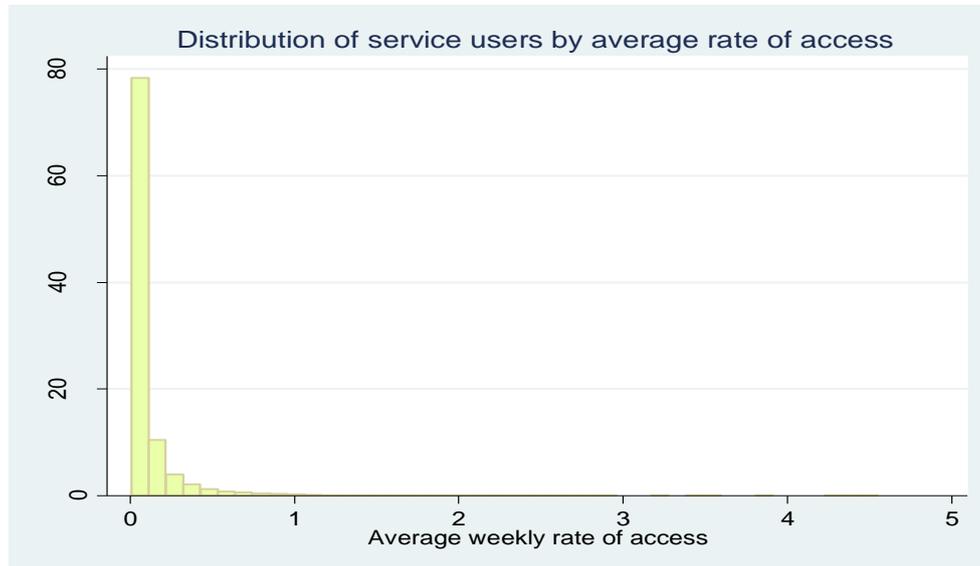
Service users accessed LLGA sessions at a mean rate of 0.11 (0.23) times a week, with a range between 0.006 and 4.55. While the vast majority did not, 309 participants (1.3%) established a habit of access to the sessions equal or greater than one time a week, on average, throughout their entire access period.

Figure 3.3 Number of service users with a rate of access of at least one time a week
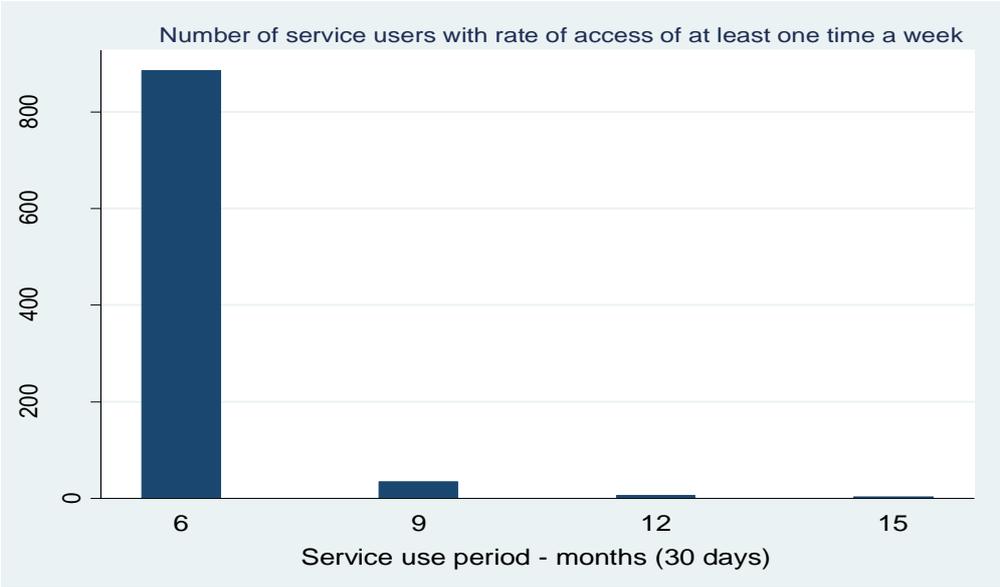
Breaking down the average rate by discrete periods, Figure 3.3 above shows the number of service users who sustained a rate of access to LLGA sessions of at least one time a week, over increasingly long periods of service use. From 876 participants over a 6-month period, the number decreased drastically to 34 over nine, 6 over 12, and 3 over 15 months beyond which no participants crossed the access threshold of 1 time a week. Comparing Figure 3.2 and 3.3, an apparent heterogeneity in the pattern of access was showed, with service users concentrating the large majority of their accesses within the first months after registration.

As shown in Table F.2 (see appendix F), RR ratios indicated that being older, particularly if over 64 years old, male and already active at baseline was positively associated with the probability of higher rate of access, with a consistent pattern across categories of outcome. Except for living in top 20% IMD score areas being associated with a 23% higher probability of accessing LLGA sessions at a rate above median value, compared to non-deprived

areas, there was no evidence to suggest that IMD status was correlated with rate of service use.

3.4.3 Facility use

Considering service use a sub-set of facility use, eight combinations of participants' use were possible. Figure 3.4 illustrates the distribution of participants by facility use status and its combinations. Using a Euler diagram, intersect zones represent sub sets that had common cases.

Figure 3.4 Distribution of participants by facility use status
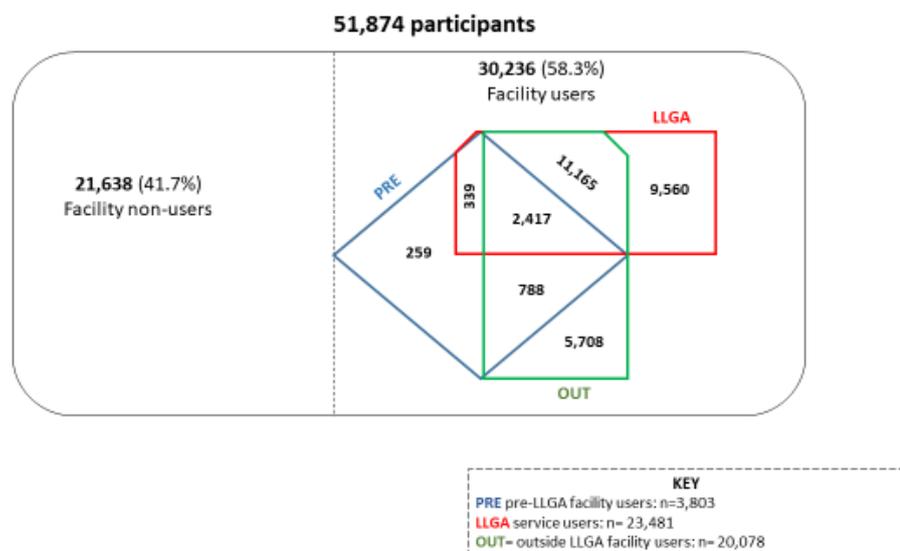
The majority of residents who signed up to the programme engaged with the leisure centre activities, at least once (58.3%), either in the six months before or during the 39 months of programme duration. Almost half (45.3%, n=23,481) used the service at least once, while 38.7% attended at least one leisure centre session outside LLGA during the programme. A 57.8% of service users accessed at least one leisure centre session outside LLGA

during the programme, while 72.5% of those who used the facilities before the programme was implemented accessed at least one LLGA session.

3.4.4 Missing data

As shown in Table 3.2, data on six of the identified variables presented missing observations in non-negligible proportions. These data were collected at baseline thorough a questionnaire that was administered only to residents registering from 01.04.2015 onward (i.e. cohort 2 participants, 37.5% n=19,438).

Considering missingness first within this sub-sample, for four variables (ethnicity, education, employment and relationship status) 32.7%, while for body mass index 66.8% of observations were missing. Missing data on body mass index could not be safely assumed to be MCAR. This is because social desirability bias is likely to occur with this and other self-report measures (Jago et al., 2007). However, this may not have been the only reason for missingness. The fact that the same proportions of observations (32.7%) were missing for the four variables could indicate issues related to data handling. This effect alone would be less of a problem, however, from a bias perspective. In fact, data handling issues could be classed as administrative censoring, whereby, the process generating missing values can be reasonably believed to be random.

Table 3.3 Differences between cohort 1 and 2

Characteristic	Category	Cohort 1 n=32,436	Cohort 2 n=19,438	p-value
Age	16-40 y	58.7%	66.1%	<0.001
	41-64 y	33.1%	28.8%	

					Notes:
	>64 y	8.2%	6.1%		
Gender	Female	62.5%	62.2%	0.460	
	Male	37.5%	37.8%		y=years
Index of multiple	Non-deprived	79.7%	81.8%	<0.001	
	Top 20%	6.7%	18.2%		.
deprivation	Top 10%	11.6%	0%		
	Top 3%	2%	0%		
Physical activity	Inactive	28.6%	29.7%	<0.001	
	Insufficiently active	35.3%	39.9%		
category	Moderately active	22.1%	20.2%		
	Active	14.0%	10.2%		
Service use status	Service users	50.1%	37.2%	<0.001	
	Service non-users	49.9%	62.8%		
PRE-LLGA	Pre-LLGA facility user	10.9%	89.1%	<0.001	
	Pre-LLGA facility non-user	1.4%	98.6%		
OUT-LLGA	Outside-LLGA facility user	48.5%	22.4%	<0.001	
	Outside-LLGA facility non-user	51.5%	77.6%		

Assuming for a moment that missing observations of those variables were due to administrative censoring only, another issue was that to assess whether those observations could be representative of the overall sample. To this purpose, table 3.3 above shows the differences between the two cohorts, in terms of observed characteristics.

Results showed that the two cohorts were statistically different, overall. What mattered was the extent to which it was plausible to believe that the observed differences did not any effect at all (MCAR scenario), or could explain the selection mechanisms for the five variables above (MAR scenario).

While a MCAR seemed not plausible, any attempt to obtain complete baseline data was deemed merely speculative. This also in light of a non-monotonic pattern of missingness occurred which complicated the issue further. Although application of multiple imputation methods or numerical simulations might have allowed dealing with such pattern (Clavel et al., 2014), the proportion of missing observations occurred, as well as the strong assumptions needed to make would have resulted in low credibility of

103

imputed values. For this reason, it was decided not to include these five variables, as well as ill health status, in the analyses. This had analytical implications in terms of loss of explanatory power of the regression models and omitted variable bias being potentially induced.

3.5   Distributional effectiveness

The sections below show the observed levels of PA before and after registration to LLGA. These are followed by the results of a distributional effectiveness analysis conducted on the sample of participants, who were obtained using the two measures of behaviour change (survey measure and card swipe measure).

3.5.1   *Before and after exposure*

Table 3.4 shows the distribution of PA categories at baseline (before exposure, "pre"), and after registration ("post"), on average and by IMD status, for the two measures.

The baseline distribution of PA categories appeared evenly balanced across IMD status, with the most deprived group being slightly more inactive then the average cohort. IMD status also seemed to play a marginal role, in that respect, with both follow-up measurements.

Table 3.4 Distribution of physical activity categories before and after registration

		Physical activity category			
		INA	INS	MOD	ACT
Baseline NAD **PRE**	AVERAGE	29.0%	37.0%	21.4%	12.6%
	NON-DEPRIVED (n=41,737, 80.5%)	28.1%	37.6%	21.7%	12.6%
	IMD TOP 20% (n=5,722, 11%)	32.6%	35.8%	20.1%	11.5%

		INA	INS	MOD	ACT
N=51,874	IMD TOP 10% (n=3,755, 7.2%)	32.8%	33.9%	20.5%	12.8%
	IMD TOP 3% (n=660, 1.3%)	34.5%	32.7%	18.5%	14.2%
Post-registration NAD **POST** n=547**	**AVERAGE**	7.3%	32.4%	42.6%	17.7%
	NON-DEPRIVED (n=461, 84.3%)	7.6%	32.7%	41.9%	17.8%
	IMD TOP 20% (n=71, 13.0%)	5.6%	32.4%	45.1%	16.9%
	IMD TOP 10% (n=15, 2.7%)	6.7%	20.0%	53.3%	20.0%
	IMD TOP 3% (n=0)	NA	NA	NA	NA
Baseline NAD + weekly rate of service use **POST** n=20,967*	**AVERAGE**	25.4%	33.9%	22.2%	18.6%
	NON-DEPRIVED (n=17,460, 81.3%)	24.3%	33.9%	23.8%	14.2%
	IMD TOP 20% (n=2,122, 10.1%)	26.8%	38.4%	22.8%	12.0%
	IMD TOP 10% (n=1,551 7.4%)	29.4%	32.9%	23.1%	14.6%
	IMD TOP 3% (n=248 1.2%)	25.4%	33.9%	22.2%	18.6%

*529 service users improved their PA category, based on change observed after 6 months after registration. ** mean follow-up time= 29.1 (15.1) weeks. Notes: INA= inactive, INS=insufficiently active, MOD=moderately active, ACT=active; IMD=index of multiple deprivation status; NAD=number of active days (questionnaire data).

Overall, a positive shift in cohort average PA level was showed, with the inactive category particularly decreasing in relative proportions with the second survey measurement. The two follow-up measurements, however, differed markedly in distributional terms, particularly with regard to the proportions of inactive and moderately active participants.

For only 547 (survey respondents) of the 51,874 participants, second questionnaire data were available. Of these, 277 (50.6%) increased their PA category, 202 did not change, while 68 reported a lower PA level. On the other hand, 20,976 participants (89.3% of service users) accessed LLGA sessions at least once within the first 6 months since their registration to the programme. Of these 20,976 service users, as already shown in Figure 3.3, 876 accessed the sessions at a weekly rate of at least one time a week. Of

these, 529, that is, 2.25% of service-users increased their PA category by accessing the programme sessions.

This difference between the two outcome measures in terms of proportion of participants that improved PA level (50.6% of survey-respondents vs 2.25% of service users) suggested that improvements in self-reported overall PA were not necessarily associated with regular attendance of LLGA sessions.

3.5.2 Interaction between service use and survey response

From a 45.3% of adults registered to the programme who accessed LLGA sessions at least once, the proportion of service users increased up to 75.9% (415 of 547) among survey respondents, suggesting a higher likelihood of response to the second survey questionnaire from this sample subgroup. However, a statistically non-significant association was found (n=547, χ^2 test, p=0.866) when testing between the probability of increasing PA category from survey response and accessing the programme sessions at least once. On the other hand, increasing PA category from weekly rate of service use was positively associated (n=385, χ^2 test, p=0.001) with the probability of reporting an increased PA category through the survey.

3.5.3 Ordered logistic models

Distributional effectiveness results are presented in Tables 3.5 and 3.6 below. These include ordered logistic regression coefficients estimated using the two outcome measures, respectively.

Table 3.5 Distributional effectiveness estimates – survey measure

n=547				Pseudo R^2=0.07
Variable	Category	β	SE	P value
PA#IMD dummy	INA NON-DEPR	reference		
	INA DEPR	0.60	0.44	0.177
	INS NON-DEPR	0.66**	0.22	0.003
	INS DEPR	0.61	0.35	0.076
	MOD NON-DEPR	**1.64**	0.26	<0.001
	MOD DEPR	**2.26**	0.53	<0.001
	ACT NON-DEPR	**2.51**	0.36	<0.001
	ACT DEPR	1.76**	0.74	0.006
Age group	16-40 y	reference		
	40-64 y	-0.01	0.17	0.953
	>64 y	0.28	0.31	0.355
Gender	Female	reference		
	Male	0.28	0.17	0.105

Notes: see Table 3.2, IMD DEPR= IMD top 20 % or top 10% score; Boldface indicates $p<0.001$; *p=<0.05;**p=<0.01

Because of a limited number of cases per category for this outcome measure, IMD variable was dichotomised in non-deprived vs deprived. A positive and fairly consistent pattern across categories of PA was found suggesting a greater effect for non-inactive individuals, relative to inactive. However, within-PA category results also showed that, except for the group of inactive, adults from deprived areas were more likely to report increases in PA category than those from non-deprived areas (Wald test, p<0.001). Neither age nor gender showed significant effects on this change.

A pattern of association similar to the previous specification was found when analysing change in PA category from the other outcome measure (see Table 3.6).

Table 3.6 Distributional effectiveness estimates – card swipe measure

n=20,967				Pseudo R^2=0.91
Variable	**Category**	**β**	**SE**	**P value**
PA#IMD status	INA NON-DEPR	reference		
	INA top 20%	-0.10	0.25	0.699
	INA top 10%	0.37	0.23	0.107
	INA top 3%	0.93*	0.48	0.051
	INS NON-DEPR	**10.2**	0.32	<0.001
	INS top 20%	**10.3**	0.38	<0.001
	INS top 10%	**10.2**	0.43	<0.001
	INS top 3%	**10.9**	0.66	<0.001
	MOD NON-DEPR	**21.0**	0.56	<0.001
	MOD top 20%	**20.5**	0.65	<0.001
	MOD top 10%	**21.2**	0.65	<0.001
	MOD top 3%	**22.7**	0.72	<0.001
	ACT NON-DEPR	46.1	800.8	0.954
	ACT top 20%	46.2	2453.4	0.985
	ACT top 10%	46.2	2636.4	0.986
	ACT top 3%	46.3	6184.3	0.994
Age group	16-40 y	reference		
	40-64 y	**0.77**	0.01	<0.001
	>64 y	**1.50**	0.12	<0.001
Gender	Female	reference		
	Male	**0.74**	0.09	<0.001

Notes: see Table 3.2

The two intermediate categories of insufficiently and moderately active appeared more likely to improve PA category, compared to the inactive group. The group of active, by design, could not improve PA. When assessing change within PA categories, individuals from the most deprived areas (i.e. IMD top 3% score) appeared to be slightly more likely to improve their baseline PA compared to non-deprived residents (Wald test, p<0.001). In this configuration, age and gender showed to play a small but statistically significant role on change in PA category.

3.5.4 Physical activity transition probabilities

In order to allow for comparison between the two outcome measures, a dichotomised version of IMD was used. Tests for proportional odds assumptions across PA categories was negative for the survey outcome model, but positive for the card swipe outcome model. Despite a series of attempts including changes to model specification and methods of categorisation, a partial proportional odds regression model could not be accepted, as producing problematic estimates (negative probabilities for part of the cases) likely to be due to lack of sufficient data points per sub-categories of the main explanatory variable. Therefore, the ordered approach was used to estimate transition probabilities between PA levels, which are shown in appendix G. This implied to assume that LLGA had an equal effect on the groups of physically non-active (inactive, insufficiently active and moderately active), therefore potentially hiding differential effects between these categories.

No statistical differences were found between deprivation groups, within the group of inactive. Having included the whole sample (n=51,874) in these analyses led to marginal probabilities of transition between PA states, whose impact will be presented in chapter n. 5, in cost-effectiveness terms.

3.6 Handling selection bias

Appendix H reports estimates from a multivariate probit model on the probability of survey response, based on participant characteristics. Except

for those reporting one active day a week at baseline, higher level of self-reported PA was not associated with the probability of providing follow-up outcome data. Being older, a service user and in particular belonging to cohort 2 was positively associated with higher probability of response, relative to their respective reference categories.

However, date of registration was showed to be, on average, negatively associated with response. This suggested that an in-person modality of data collection and the use of the IPAQ questionnaire were more effective, in this regard, than a web or email-based approaches and the use of the lifestyle questionnaire.

3.6.1 Selection mechanisms

Table 3.7 below compares the regression coefficients estimated using the base-case approach (last observation carried forward) with those estimated in the three other selection scenarios. Deviation from the base-case approach, in terms of average change in NAD, is showed in column 6.

Relative to the base-case approach, all the three selection scenario results were more favourable, in terms of change in PA associated with LLGA, by a large margin. The complete case approach generated a deviance from a last observation carried forward approach (MCAR) of more than one full active day a week per participant. The other two selection models adjusted these results to divergent directions. Adjusting for unobserved heterogeneity (MNAR) led to results that were 8.6% more favourable than the MCAR scenario. The most

conservative approach showed to be that of MAR, although being by only 10.9% lower than MCAR results.

Table 3.7 Selection scenario results

Approach	Effectiveness coefficient	Age 40-64 y	Age >64 y	Gender	IMD status	Deviation from ITT coefficient (NAD)
LOCF	0.012	**-0.145**	**0.281**	**0.313**	-0.111	reference
	(0.012)	(0.024)	(0.013)	(0.012)	(0.015)	
MCAR	**1.068**	**-0.149**	**0.281**	**0.313**	-0.109	(+ 1.056)
	(0.082)	(0.018)	(0.034)	(0.017)	(0.021)	
MAR	**0.953**	**-0.058**	**0.212**	**0.268**	-0.148	(+ 0.941)
	(0.172)	(0.129)	(0.050)	(0.027)	(0.032)	
MNAR	1.159**	-0.149	0.343	0.312**	-0.131	(+ 1.147)
	(0.398)	(0.018)	(0.221)	(0.113)	(0.139)	

Notes: LOCF=last observation carried forward; see Table 3.2

In order to covert the regression coefficients estimated under MAR and MNAR assumptions into PA transition probabilities, the following formulas were used:

$$P_m{}^{MAR} = P_m{}^{MCAR} * [1+ (\Delta\overline{NAD}_{MAR} - \Delta\overline{NAD}_{MCAR}) / \Delta\overline{NAD}_{MCAR})]$$

$$P_m{}^{MNAR} = P_m{}^{MCAR} * [1+ (\Delta\overline{NAD}_{MNAR} - \Delta\overline{NAD}_{MCAR}) / \Delta\overline{NAD}_{MCAR})]$$

In other words, the change in probability of moving from a baseline PA category to any of the other three categories is proportional to the difference in average NAD estimated using a complete case analysis approach.

3.7 Chapter summary

This chapter focussed on describing the case study data, the estimation of distributional effectiveness of LLGA programme and handling of selection bias.

LLGA was able to attract a large number of adult residents. Around a quarter of participants belonged to target groups of physically inactive and living in deprived neighbourhoods, and almost half attended at least one programme session. The programme was associated with a positive distributional effect favouring the more deprived groups, although by a small margin, according to the card swipe measure. Conversely, an inconsistent pattern of effect was found when analysing the survey measure, with an average negative effect on PA inequalities. There appeared to be a positive association between improvement in PA category from the self-report measure and improvement due to a sustained rate of service access.

Two measures of behaviour change (i.e. change in overall PA behaviour over time) were identified within the case study data to analyse transitions between the four PA categories. The extent of missing information and baseline missing data limited the ability to reliably estimating the distributional effectiveness of the programme. To enable an EE of LLGA in Chapter 5, programme effect on change in PA was obtained applying a pragmatic approach (i.e. last observation carried forward), to reduce the risk of overestimation.

Coefficients obtained from the two specified ordered logistic models (based on the two outcome measures) were converted into transition probabilities between PA categories. The effectiveness parameters estimated from analysis of the card swipe measure will populate the developed decision analytic model (Chapter 4) for EE of LLGA programme (Chapter 5). The effectiveness parameters estimated from analysis of the survey measure will instead be used as a benchmark when testing the assumption regarding the causes of survey non-response in chapter 6.

The issue of selection bias was addressed formally by simulation of three alternative scenarios regarding the reasons for survey follow-up outcome data to be missing (MCAR, MAR and MNAR). All three scenarios produced effectiveness estimates which were more favourable than the pragmatic approach and deviated from one another by a small margin. These adjusted estimates were also converted into transition probabilities that will be used to explore the impact of the respective mechanisms of missingness assumptions in chapter 6.

4 Chapter

Modelling of impact

4.1 Chapter outline

This chapter presents a modelling approach devised to address the two modelling shortcomings of population-level impact (i.e. ability of the model to capture heterogeneous and equity-relevant intervention effects across different groups of individuals within a given population distribution) and behaviour change maintenance over time (i.e. formal extrapolation of decay in intervention effects), which were identified in the review (chapter 2).

The aim was to develop a general and simple modelling approach that was able to address these two key shortcomings and could be further adapted to other evaluation scenarios. Based on this modelling approach, a decision-analytic model is developed to assess the cost-effectiveness and health inequality impact of universal programmes to promote PA in the adult general population. To also enable an EE of the LLGA programme (chapter 5), the decision-analytic model structure is adapted, to be populated with parameters estimated from analysis of the case study data (chapter 3).

The conceptual design and mechanics of the proposed modelling approach are first described in sections 4.2. Section 4.3 presents the developed decision-analytic model structure, parameters and sources, and validation procedures followed. The last section 4.4 summarises the chapter's key messages and implications.

114

4.2 The modelling approach

The growing use of EE to inform decision-making in the health sector has seen an increased importance of decision-analytic modelling as a means of evaluation (Briggs et al. 2006). Especially in public health settings, all the relevant information and evidence needed to make an investment decision will rarely be possible to assess by a RCT. In particular, comparison of all the relevant options, linking intermediate to final outcomes and extrapolate cost and effectiveness beyond trial data is often needed to perform an EE. A framework able to do so in a context of uncertainty is therefore needed. While its role has been controversial (Buxton et al., 1997), decision-analytic modelling has seen an increased importance over the last two decades. Evidence of this is the recommendation of its use within the National Institute for Health and Care Excellence for public health evaluation methods guidance (National Institute for Health and Care Excellence, 2012).

A number of decision-analytic modelling approaches can be used to model the impact of health promotion interventions (Briggs et al., 2016). Choice of approach is dependent on the stated decision problem and boundaries of the model. Among cohort-level approaches, Markov models are the most common form of models used in decision analysis for economic evaluation (Briggs et al., 2006 book). Through a Markov approach, intervention options can be modelled over time by representing the possible consequences in discrete states. Transitions between Markov states over a series of discrete time periods (cycles) are allowed for representing stochastic processes, such as good health to disease progression (Briggs et al., 2006).

4.2.1 Conceptual design and mechanics

To model PA, a compositional modelling approach was devised. Here, compositional refers to the modular way of conceptualizing Markov states and their dynamics, which has seen application in human genetics (Blossey et al., 2006) and medical research (Ma et al., 2018).

Instead of considering Markov states as single independent entities, they are considered as sub-parts of one macro-level state, which interact with one another, and are modelled jointly. Therefore, a change in proportion of any sub-part generates an opposite-sign change in proportion of the other sub-parts. Let us consider a simple three-state MC, as shown below in Figure 4.1. And let us focus on the part of the model that represents the focus of prevention initiatives: preventing healthy individuals from becoming ill (i.e. P_1).

Figure 4.1 Three-state Markov chain
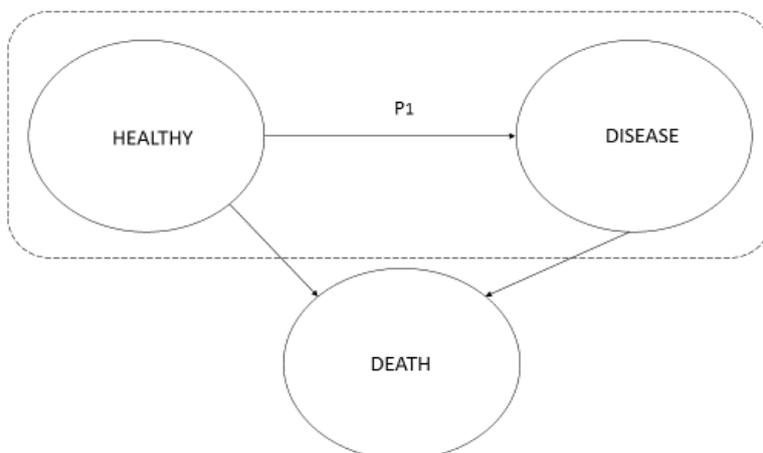
The Healthy state can be seen as the population composition of PA habits (i.e. frequency distribution of PA states), which can be impacted by an intervention at a certain point in time.

Through a compositional approach, the Healthy state is no longer a single Markov state, but a composite entity (macro state) whose behavior is dependent upon the frequency distribution of inner micro-states (i.e. PA levels). In this case, the behavior in question is the probability to transit from a Healthy to a Disease state (P1).

In modelling terms, the Healthy state can be modelled as an Embedded MC (EMC), also known as Nested MC or Embedded Jump Chain (Douc, 2018). EMC methods, an extension of discrete-time MCs, have been applied in many fields to capture complex system-level behaviors (Tagliaferri et al., 2016).

In accordance with the current UK PA recommendations for adults (Health, 2011a), and in line with the approach used to assess the distributional effectiveness of the LLGA programme (chapter 3), four PA levels (micro-states) were defined: inactive = zero; insufficiently active = 1 or 2; moderately active= 3 or 4; active = at least 5 active days a week.

Graphically, the EMC can thus be represented as displayed in Figure 4.2 below.

Figure 4.2 Nested Markov chain

HEALTHY
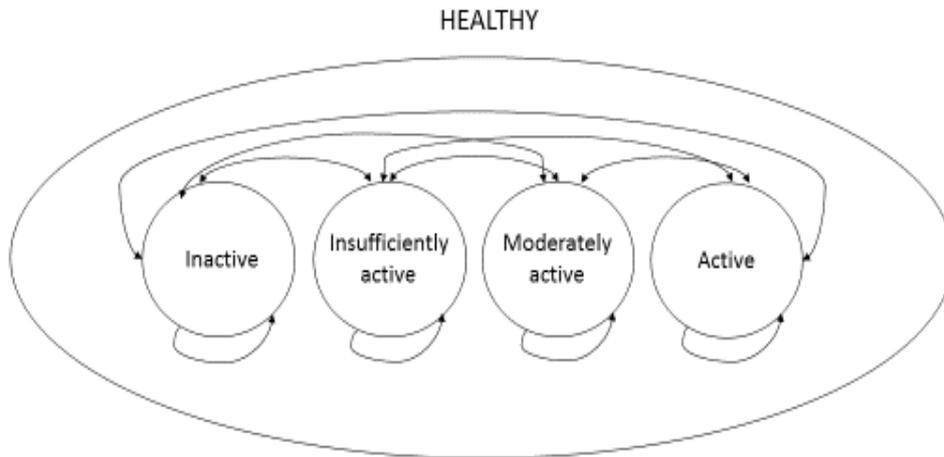
Within the proposed EMC, the full range of possible transitions (16) between the same (4) or different (12) PA levels is allowed to occur over time. This feature extends existing models, to fully capture baseline PA related heterogeneity of intervention effects.

4.2.2 *Natural course of physical activity*

To model time-dependent dynamics of PA, a continuous-time MC approach was integrated into the EMC (Mhoon et al., 2010). Through a continuous-time MC change in probability of transition between Healthy and Disease states (P1) over time was made depending upon time-dependent changes in composition of the Healthy state (i.e. frequency distribution of 4 PA levels).

 To illustrate, consider a closed cohort of individuals grouped according the four PA levels at baseline. In absence of intervention, P1 will be equal to the weighted (by group proportion) probability of the four PA levels to transit from Healthy to Disease. In mathematical notation terms, given $p1_i$ = probability of PA state i to transit from Healthy to Disease and w_i = proportion of individuals in PA state i relative to the sum of the 4 PA states:

$$P1 = \frac{\sum_{i=1}^{4}(p1_i \, x \, w_i)}{\sum_{i=1}^{4} w_i} \qquad \text{with } \sum_{i=1}^{4} w_i \text{ always adding up to 1.}$$

Therefore, the probability P1, which is conditional on the frequency distribution of the four PA states (i.e. population-level proportions of PA levels), will change "naturally" over time due to the different risks of disease across PA levels. If $p1_i$ are assumed as constant, the probability distribution of this process is not conditional on time. In other words, Healthy to Disease is a stationary process.

4.2.2.1 Embedded Markov Chain

Given the EMC structure, any time-discrete step can be described by a square matrix. The transition probabilities between PA levels ($P_{t-1,t}^{nat}$) can be represented as a square matrix. The values on the diagonal represent the four transitions from and to the same state, while the off-diagonal cells include the transition probabilities (TPs) between different PA states. A zero change in PA level will mean $P_{t-1,t}^{nat}$ to be an identity matrix (*I*), that is, all diagonal values are equal to one.

For each of the four PA states, for example the inactive state, there is only one possible transition from and to the same state (i.e. TP_{11}) and three possible transitions to the three other levels (i.e. TP_{12}, TP_{13} and TP_{14}). These two types of transitions can be analyzed as two complementary events, whose combined probability ($P_{r \mid} P_m$) must equal to 1: to remain in the same state ($P_r = TP_{11}$) or to move to another state ($P_m = TP_{12} \mid TP_{13} \mid TP_{14}$).

Figure 4.3 below provides a graphical representation of the two events for the inactive state.

Figure 4.3 Remaining in or moving to another physical activity state
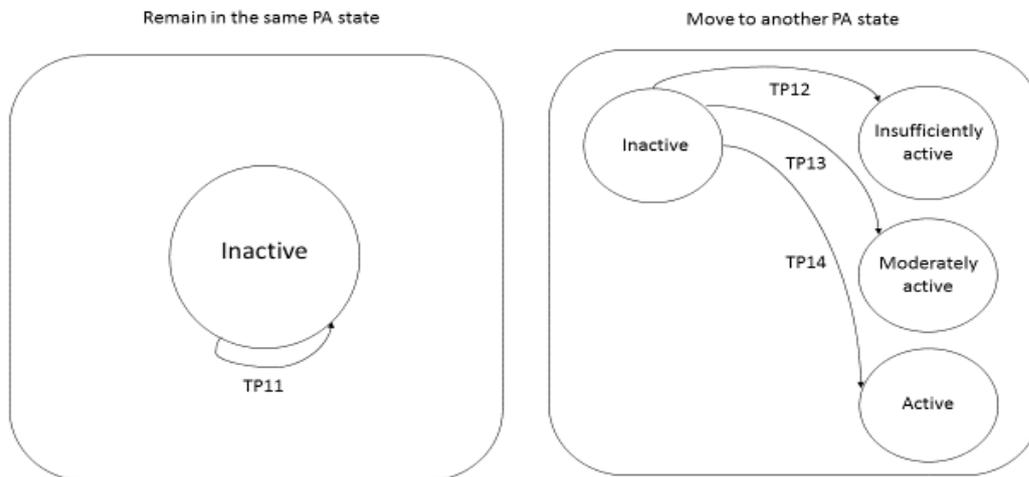
4.2.3 Population-level impact

From a modelling perspective, an intervention effect represents a shift from a "natural" stationary process, whereby the probabilities of transition between PA states move away from their natural course.

In terms of the two complementary events described above, using an example for the inactive state, an intervention effect can be represented as the *difference* in probability of the second event P_m (i.e. $TP_{12} + TP_{13} + TP_{14}$), relative to $P_{t-1,t}^n$. As being modelled jointly, however, any change in P_r will be paired with an opposite-sign change in P_m (i.e. all rows of the PA transition matrix must sum up to one).

4.2.3.1   Differential rates of transition

In matrix terms, given a matrix $A_{t-1,t}$, namely, the intervention effect matrix for the period t-1 to t, a PA state space S = [1,2,3,4], and $P_{i,j}$ = probability that the chain will move to state j, given that is in now in state i, then:

$$A_{t-1,t} = \begin{bmatrix} -a_1 & a_1P_{1,2} & a_1P_{1,3} & a_1P_{1,4} \\ a_2P_{2,1} & -a_2 & a_2P_{2,3} & a_2P_{2,4} \\ a_3P_{3,1} & a_3P_{3,2} & -a_3 & a_3P_{3,4} \\ a_4P_{4,1} & a_4P_{4,2} & a_4P_{4,3} & -a_4 \end{bmatrix}$$

where a_i = differential (from $P_{nat}^{t-1,t}$) transition probability which is expressed as a rate of transition out of state i. The term $a_i P_{i,j}$ can be interpreted as the differential rate of transition between different PA states, under the condition that $a_i P_m = - a_i P_r$. If a_i is constant, a discrete-time chain is represented.

In regression terms, a_i may represent a treatment effect (coefficient) estimated from a (generalised) ordered logit model. Given a baseline composition of PA states represented by the vector θ^{t-1}, in order to obtain the post-intervention θ^t, a matrix multiplication is simply needed:

$$\theta^t = \theta^{t-1} \times P_{t-1,t}^{nat} \times A_{t-1,t}$$

Thus, the model considers both baseline PA-related heterogeneity of intervention effect, as well as population distributions, in terms of proportions of individuals within the four PA levels. Therefore, differential impacts of any universal intervention on existing *health inequality* between population sub-groups (e.g. socio-economic sub-groups) can also be captured, by simply replicating the model structure for *each* of the sub-groups that reflect the inequalities of interest.

4.2.4 Behaviour change maintenance

The other shortcoming identified in the systematic review (see Chapter 2)

was that of maintenance of behavior change over time. Individuals impacted

by the intervention will be likely to converge to their natural course of PA

habits, at a certain rate (Van Dyck et al., 2017). In other words, any causal

effect on PA behaviors will not be likely to remain constant over time (i.e. $A_{t-1,t}$, is not an identity matrix) but to have a rebound trajectory which, in turn, is

likely to be dependent upon baseline PA level.

4.2.4.1 A discrete-time survival approach

These rebound trajectories can be broken down by time periods

(corresponding to the Markov cycles' length) and analysed, provided relevant

data are available, through a discrete-time survival approach. The idea is to

conceive maintenance of behavior change as a survival function, whereby

survival equals the residual intervention effect at a certain point in time. A

fictitious representation of exponential rebound effect is in Figure 4.4.

Figure 4.4 Illustrative example of maintenance of behaviour change over time

Maintenance of behavior change over time
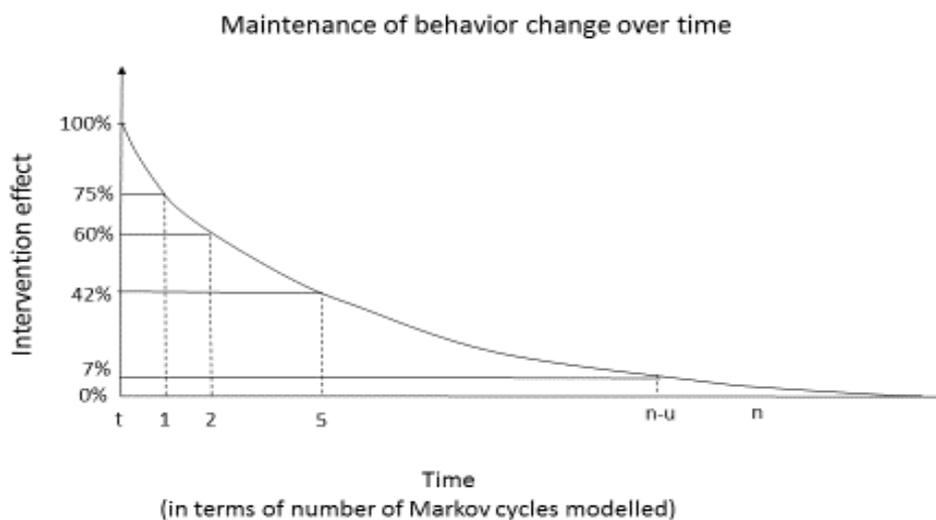
Residual intervention effect is represented on the y axis and Markov cycle number on the x-axis. From time t (post-intervention, 100%), the intervention starts converging gradually towards zero value over time. In this example, at the beginning of the first cycle after intervention, 25% of intervention effect is decayed. In other words, the residual intervention effect is at 75% of its original magnitude ($P_{t,t+1}^{res}$=0.75), In the next cycle (cycle 2), 40% of the initial programme effect is faded out ($P_{t,t+2}^{res}$=0.60), corresponding to a 20% loss of intervention effect from the previous cycle (15/75), and so on.

4.2.4.2   Residual intervention effect

To compute the residual intervention effect (p_i^{res}) for a given cycle, three steps need to be followed. First, a survival model needs to be specified. In order to allow for extrapolation of effects over time, in line with (Briggs, 2006), a parametric approach needs to be chosen. Choice of distribution can be informed through testing of alternative distributional forms and comparison made in terms of Akaike or Bayesian information criteria.

Using the notation in section 4.2.2.1:

$A_{t-1,t}$ = intervention effect matrix (100% of residual effect),

then:

$P_{t,t+u}^{res}$ = residual intervention matrix at cycle t+u

Once calculated the rates of decay (λ_i) between each of the considered cycles, these can be converted using the following formula (Briggs, 2006):

$$p\lambda_i (t,t+1) = 1-\exp[\lambda_i(t)^y - \lambda_i(t+1)^y]$$

with $p\lambda_i (t,t+1)$ =probability of effect decay of effect from the previous cycle.

To calculate the probability of residual effect:

$$p_i^{res}(t,t+1) = 1 - p\lambda_i(t,t+1)$$

The general formulation for calculating the the probability of residual intervention effect left up to time t+u:

$$P_{t,t+u}^{res} = \prod_{t=0}^{u}(p\lambda_i\ (t, t+u))$$

A simple matrix multiplication of the cycle probabilities of residual intervention effect from time zero (i.e. post-intervention) up to the cycle t is therefore needed. Using the example in Figure 4.4, once $P_{t,t+u}^{res}$ is computed, for example, for u=5 (i.e. $P_{t,5}^{res}$ = 0.42), the residual intervention effect matrix for cycle 5 is obtained by multiplying the intervention effect matrix ($A_{t-1,t}$) by the *each* of the respective cycle probability of residual effect (p_i^{res}) up to cycle 5.

From the previous step, a series of subsequent PA transition probability matrices describing the progressive loss of intervention effect over time can be computed. In implementation terms, to adjust for the respective loss of intervention effect at cycle n:

$$\theta^n = P(_{0,n-1}^{res})^{-1} \times P(_{0,n}^{res}) \times \theta^{n-1}$$

Using the example in Figure 4.4:

For cycle 1 $[A_{t-1,t}^{-1} \times P(_{0,1}^{res})]$ $= -100\%\ \text{effect} + 75\%\ \text{effect} \times \theta^1$

For cycle 2 $[P(_{0,1}^{res})]^{-1} \times P(_{0,2}^{res}) \times \theta^2$ $= -75\%\ \text{effect} + 60\%\ \text{effect} \times \theta^2$

For cycle 3 $[P(_{0,2}^{res})]^{-1} \times P(_{0,3}^{res}) \times \theta^3$ $= -60\%\ \text{effect} + 42\%\ \text{effect} \times \theta^3$

To compute the rebound at a given each cycle, the intervention effect left from the previous cycle is first subtracted (through matrix inversion), to be then replaced by the current cycle's residual intervention effect (through matrix multiplication).

4.3 The decision-analytic model

4.3.1 Model structure

The decision-analytic model general structure is presented below (see Figure 4.5). The model features the continuous-time MC developed and described above for the Healthy-Disease transition, together with two discrete-time MCs from Healthy and from Disease states to Death.

Figure 4.5 Decision-analytic model diagram
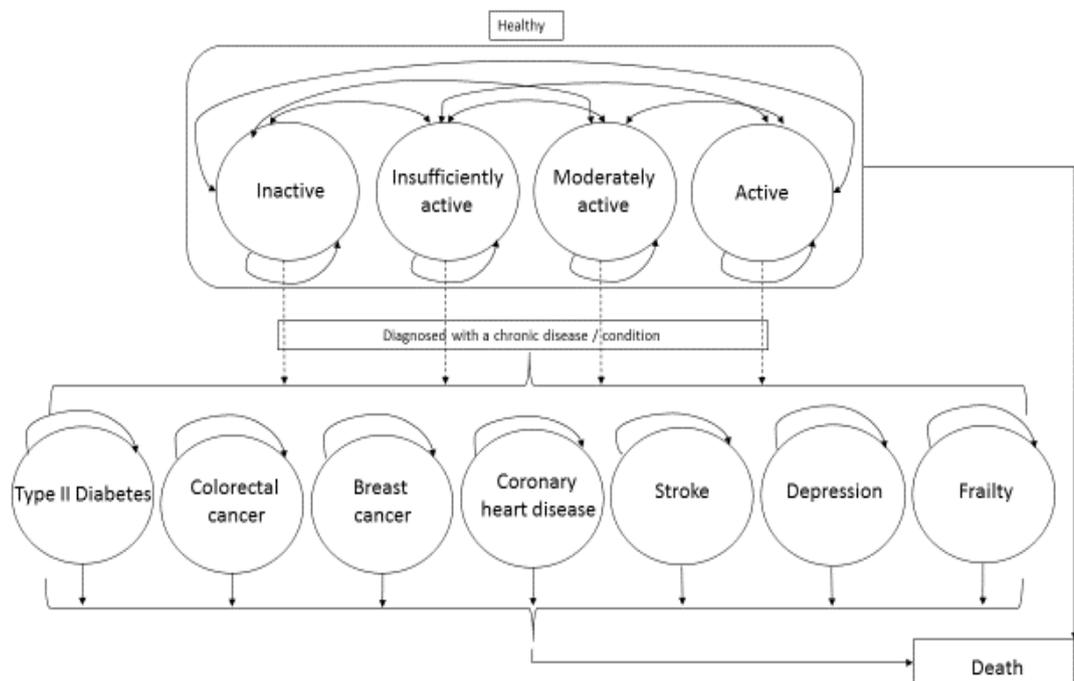
Healthy subjects, defined as not being diagnosed with any chronic disease / condition, may develop only one of the seven identified diseases / conditions. Healthy subjects can: either progress to Death at an age-dependent all-cause mortality rate, or progress to any of the disease states. Once entered a disease state, PA level no longer affects health-related quality of life. Cohort members can either remain in a disease state or move to the absorbing state (i.e. Death), at an increased RR compared to Healthy members.

In line with previous research (Asaria et al., 2016a, Love-Koh et al., 2015), neighbourhood-level deprivation was identified as the equity-relevant characteristic (i.e. IMD status) to consider for modelling distributional impacts. PA and IMD status were considered as independent contributors to risk of disease, but neutral to mortality risk. In order to characterise cohort's baseline (Healthy macro-state) health-related quality of life, micro-state specific health utility values were assigned (see section 4.3.4 for sources of parameters).

The model is designed to project intervention effects of any universal strategy on a closed cohort of individuals over a lifetime, (i.e. until members reached 100 years), with simulations that can be stratified by age and gender. Model outputs include: number of cases averted (i.e. number of diseases and deaths), LYs, QALYs and costs saved. Disease-specific health utility decrements are applied when members progress to any disease state. The model was built in Excel (Microsoft Office 2016).

4.3.2 Chronic conditions included in the model

Regular PA has been associated with reduced risks of many chronic conditions (World Health Organization, 2008)). Disease identification was informed by the last scientific report from the Physical Activity Guidelines Advisory Committee (2018), which forms the basis of the current UK PA guidelines (Chief Medical Officers, 2019). This report assessed the relevant available evidence from systematic reviews and meta-analysis on the PA-disease relationships against five criteria (i.e. applicability, generalisability, risk of bias or study limitations, quantity and consistency and magnitude and precision of effect), grading it as either: strong, moderate or limited. Only chronic diseases and conditions for which strong evidence existed associating regular PA with lower health risk were selected, in line with global burden of disease studies (Lee et al., 2012; Ding Ding et al., 2016). These are type II diabetes, coronary heart disease, stroke, colorectal cancer, breast cancer, depression and frailty syndrome.

In line with the previous models, obesity status was not included as an intermediate variable in the pathway to disease. This is because of the complex relationship with PA and potential double-counting of costs and health outcomes (Roux et al., 2008). For the same reason, pre-clinical conditions on the pathway to chronic diseases for which strong evidence exists, such as hypertension and metabolic syndrome, were excluded. Considering the epidemiological evidence reviewed, for breast cancer, only incidence in female individuals was taken into account, while for frailty syndrome the probability to transition started when the cohort age was 65 years.

4.3.3 Adaptations to model structure for LLGA

The general model presented above was simplified. This was because of lack of data on the entire IMD distribution (i.e. only first quintile or not first IMD quintile available) and of reliable information regarding the trajectory of behaviour change following the intervention. In order to enable an EE of LLGA and make the most of the data available, heterogeneity of impact and decay of effectiveness over time were considered in terms of PA (four levels) and IMD status (two levels; IMD non-deprived, below first quintile of IMD score; IMD deprived, first quintile IMD score, eight combinations). One intervention arm (i.e. LLGA programme) and one control arm (i.e. no-LLGA) were developed. Each of these two arms were divided into two sub-groups, IMD non-deprived and IMD deprived, for each of which a model was developed. Appendix G reports the transition probabilities used to assess the cost-effectiveness of LLGA, under base-case assumptions.

For the two LLGA sub-arms, the intervention could impact the composition of PA levels within the first cycle which, was assumed to last 6 months (as the average follow-up period). The remaining period of programme duration (33 months), was divided in equal periods of 3 months each (11 cycles). All transition probabilities in the model were computed to fit the cycle length by exponential rate-to-probability functions (Briggs, 2006) .

Considering the evaluation context, no intervention effect was presumed to last longer than the programme duration. Three options regarding maintenance of behaviour change over time were developed and tested in chapter 6: no decay of effect, immediate rebound and exponential rebound.

With a no decay, maintenance is assumed to be constant at a 100% over time. An immediate rebound means that in the first cycle post-intervention (i.e. at the beginning of the 7th month since exposure), no intervention effect has remained and healthy cohort members have returned to baseline levels of PA. An exponential rebound trajectory could instead represent a gradual return to baseline homeostasis (Sport England, 2012).

While none of the three underlying hypotheses can be strongly supported by evidence, a no decay scenario seems the least likely to occur. An immediate rebound assumption, on the other hand, might have been too strict. Cohort members could react variedly to the intervention, with few sustaining their change in PA behaviour beyond the first six months (e.g. by becoming a regular gym member). From a theoretical standpoint, this could be explained by the fact that healthy members can be at different stages of change (Mhoon et al., 2010).

Through such perspective, it is thus reasonable to assume that, while a majority of participants initially impacted by the intervention will return to their baseline PA levels immediately after, part will return gradually to it. Such distribution can be represented by an exponential function (e.g. Figure 4.4). For the purpose of simulating what could have happened to the frequency distribution of *overall* PA levels beyond the first six months, an exponential model was fitted to the card swipe data using the approach described in section 4.5.

Following AIC / BIC testing, a Weibull distribution was chosen. In terms of survival analysis, card swipe data were analysed to obtain the drop-off patterns of service use (service users no longer attending the programme sessions, Figure 3.1). The origin was set to the individual date of registration to LLGA, with the "failure event" being represented by the last access to LLGA sessions Stata software version 14 was used for analysis. Wald tests were used to assess whether there was evidence of differential rebound trajectories (i.e. lambda values) between PA / IMD categories.

4.3.4 Model parameters

A literature search was conducted in Medline electronic database (via Ovid) using a combination of search terms (Appendix P) to obtain the baseline parameters to populate the developed decision-analytic model, including the reviewed economic models. These parameters were: disease probabilities and costs, relative risks of mortality, PA and socio-economic deprivation gradients relating to the selected diseases. Effectiveness parameters were sourced from the effectiveness analysis of LLGA case study (see next section). A combination of keywords relating to the identified concepts (disease names e.g. "stroke", PA, socio-economic status, cost) and search filters were applied as needed, according to three criteria:

- relevance: studies focussing on the selected disease and assessing the relationships with PA and neighbourhood-deprivation status on healthy individuals. Thus, research studies conducted on clinical populations were excluded.

- hierarchy of evidence: based on the establishes hierarchy of study design (Reviews, 2019). If systematic reviews and meta-analyses were not available, lower rank study designs were selected, in decreasing order: RCTs, cohort studies, case-control studies, cross-sectional survey and case reports.

- scope: given that the simulation of the intervention impact would be performed on a UK population, empirical studies focused on the UK population and conducted no more than 10 years prior were identified. If not available, studies based in, in order: European countries, Western countries, any country were otherwise chosen.

When no data on gradients (i.e. RRs) were available, in line with previous studies (Frew et al., 2014, Roux et al., 2008), a linear interpolation method was used. Specifically, a proportional dose-response relationship was assumed between energy expenditure rates (MET) corresponding to the four PA levels and RRs.

In line with reviewed models, for coronary heart disease and stroke diseases, a transitional tunnel state approach was used to capture the increased healthcare costs associated with the first year of experiencing the event, compared to subsequent years. Estimates of baseline health utility values for the eight PA/neighbourhood-level deprivation micro-states were obtained through regression analysis of UK national survey data (Health Survey for England), following the approached adopted by (Maheswaran et al., 2013).

4.3.5 Case study parameters

Effectiveness (i.e. intervention effectiveness matrix) and behavior change maintenance parameters (i.e. residual intervention effect matrices) were sourced from the LLGA data sets. As mentioned before, both measures of behavior change (survey and card swipe measure, see chapter 3.3.4.1) were used to populate the model.

As mentioned above, effectiveness estimates from the card swipe measure are used for an EE of the LLGA programme (chapter 5). Those obtained from analysis of the survey measure are instead used in chapter 6 to test the missingness scenario assumptions, where also all three assumptions regarding maintenance of behavior change are tested. Programme unit costs were derived from the data included in the financial audit reports (see appendix I).

4.3.6 Model validation

Good practice guidance for model validation was followed (Philips et al., 2006, Vemer et al., 2016). A number of "check alerts" were included and internal testing was conducted to test whether the model produced logical and expected outputs, respectively. As for the former, these included checking whether:

- The sum of proportions of PA levels was equal to 1 in each cycle
- The sum of proportion of members in the macro states was equal to 1 in each cycle
- Spreadsheet cells did not contain negative values

- Proportion of members in the absorbing state was equal between the intervention arms at the end of the first cycle.

Null and extreme input values were used to test the following propositions:

1. If baseline composition of PA levels was equal and intervention effect = 0 (identity matrix): no difference between Intervention and No-intervention arms.

2. If intervention effect size was positive and larger for IMD non-deprived than for the IMD deprived group: higher number of cases averted and LYs saved for IMD non-deprived relative to deprived.

3. If 2 and if health utility values = 1: larger gain in QALY for IMD non-deprived relative to deprived.

4. If baseline composition of PA levels was equal and probability of transition to a disease = 0: no difference between arms in terms of cases averted for that disease.

5. If baseline composition of PA levels was equal and probability of transition to all the diseases = 0: no difference between arms in terms of cases averted, LYs and costs.

6. If baseline composition of PA levels was equal, RRs=1 and QALY values=1: no difference between arms in terms of cases averted, LYs, QALYs and costs.

An external modeller not involved in the development of the model performed an independent review of the model's logical soundness, and tested it to ensure that it behaved in accordance with the conceptual model. After having explained the purpose, mechanics and features of the model, this review was

performed in ninety-minute long meeting by letting the external modeller check the Excel tool for errors and inconsistencies. Model outputs were also compared with results obtained using an off-the-shelf tool (Sport England, 2012).

4.4 Chapter summary

This chapter has presented a modelling solution to overcome two modelling shortcomings of previous models, which were identified in chapter 2, namely, population-level impact and behaviour change maintenance over time. The proposed modelling approach is novel and offers a flexible structure which can be adapted for EE of any universal strategy to promote healthy behaviours. The developed decision-analytic model can be used to assess the cost-effectiveness and health inequality impact of universal programmes to promote PA in the adult general population. Due to its simplicity, the model has potential for widespread application in public health settings, as being easy to understand for a lay audience of public health decision makers to which the model was targeted.

Unlike the previous models, the proposed modelling approach allows for full interaction between four PA categories which are aligned with the current UK PA recommendations for adults. Instead of relying on a structural assumption of PA being a fixed characteristic (state) unless an intervention occurs, natural courses of PA can be modelled formally by means of a compositional approach. This may be especially important when modelling PA over short periods of time or sensitive life phases (e.g. developmental age, retirement),

that is when natural fluctuations are more likely to occur (Van Dyck et al., 2017).

The developed decision-analytic model can be adapted to evaluate any intervention aimed at promoting change, as well as maintenance of active behaviours. Improvements in population health outcomes can propagated as a result of either cohort members increasing PA levels (like in previous models) or by increased probabilities of remaining in higher PA categories relative to a natural tendency towards lower PA states (e.g. during sensitive life phases). Moreover, negative intervention effects can be also formally taken into account (e.g. transitions to lower PA categories due to intervention, for example, current exercisers that are deterred by overcrowded gyms or injuries).

The model was designed to address policy concerns related to the distributional impacts of universal strategies. The developed analytical tool is able to generate sub-group cost-effectiveness estimates to inform decision-makers about the number or proportion of physically inactive adults, by neighbourhood-deprivation status, needed for any intervention to be cost-neutral. Gross and net inequality impacts of any universal PA intervention can be also generated for DCEAs to be conducted, as being done for the evaluation of the LLGA case study, using the framework for equity trade-off analysis by Cookson et al. (2017).

However, a number of limitations need to be acknowledged. Belonging to the Markov family, the proposed state-transition model shares its features and limitations (Douc, 2018). Specifically, the memoryless property with future states being dependent only on the present state. In the context of behaviour analysis, however, this represents a limitation because behaviours are likely not to be independent from past experience. In fact, the proposed model also assumes that the members within a given PA category are a homogeneous group. To this respect, an individual-level approach, such as that employed by (Gc et al., 2018) may more adequately capture that aspect of heterogeneity, provided relevant data are available.

While it explicitly incorporated concerns for health equity, the model was designed with only two levels of socio-economic deprivation, namely, IMD first quintile and the rest of IMD distribution, so as to enable the EE of LLGA. However, decision-makers may be interested in assessing the impact on the full quintile distribution, as conducted in previous studies (Asaria et al., 2016a, Dawkins et al., 2018). While the proposed modelling approach can be applied to other decision problems, impact of the intervention on individuals aged below 16 years old cannot be estimated by the developed decision-analytic model. For this population, different prevention pathways and modelling approaches would be needed. Given that the purpose of the thesis was to illustrate the implications of methodological assumptions of current models of PA, this aspect was not addressed.

Overestimations of economic benefits may derive from a lack of proper inclusion of time lags between changes in PA and disease occurrence. This was due to a lack of clear evidence on these relationships, which are likely to be age-dependent, as well as, dependent of past habits. These aspects were not addressed in the development of the model, except for frailty syndrome which starts when the cohort age is 65 years old. In addition, and in line with previous models, the model was designed in a way that made disease risks compete with one another. In reality, however, this may not be necessarily the case (Giovannucci et al., 2010). This is because the selected conditions share PA as one of their determinant factors. For instance, a reduction in risk of type II diabetes, due to an increase in PA, would generate a reduction in risk of stroke. While being structurally inexact, a competing risk mechanisms is likely to result in underestimations of the impact on an intervention, counterbalancing the risk of overestimation derived from other structural assumptions of the model.

Disease recurrence, increased health expenditure from extended life expectancy and adverse events, such as injuries from increased PA, were not formally taken into account. As for the latter, however, the compositional structure of the model allows for negative impacts of the intervention (transitions to lower PA states), with potential for capturing adverse effects. Moreover, average disease costs were assumed to be constant over time for five of the seven conditions considered in the model (type II diabetes, depression, frailty, colorectal and breast cancer).

For stroke and coronary heart disease, different disease costs for the first year, which tend to be higher than those for the subsequent periods after the event (Luengo-Fernandez et al., 2012), were specified. Disease costs for the six conditions already modelled within existing models (all except frailty syndrome) were sourced from the respective published peer-reviewed articles, which were only in part based on meta-analysis results. While possibly leading to inaccurate estimations of incremental effects, this allowed to increase comparability with the reviewed models and their results. Balancing time and resource constraints, the complexities described above were not addressed, and further work should be undertaken to overcome these shortcomings.

Finally, certain caveats must also be borne in mind. The proposed framework is based not only on a cause-effect relationship between intervention and change in PA behaviour. It also relies on a strict assumption of causality between changes in behaviour and changes in disease and mortality risks. These aspects are common to the vast majority of health promotion models, which typically rely on observational evidence. Moreover, validity of incremental estimates is pre-conditioned by the extent to which compensatory effects occur. For instance, health risk behaviours, such as excessive alcohol and smoking-related behaviours, could occur as substitution effects on the causal chain to health improvement, hence altering the impact. Addressing these concerns was deemed not achievable within the time and resources available, but they represent important aspects to

consider for development of a broader general model for promotion of

healthy behaviours, which does not exist at this time.

The decision-analytic model has been adapted for EE of LLGA programme,

of which cost-effectiveness results and methods used for uncertainty

assessment are presented in the next chapter.

5 Chapter

Economic evaluation of the LLGA case study

5.1 Chapter outline

This chapter focuses on the EE of the LLGA programme. Section 5.2 describes the methods used for assessment of the cost-effectiveness of the intervention. Section 5.3 reports the deterministic cost-effectiveness results, and the results of a threshold analysis (Rodriguez-Martinez et al., 2018). In section 5.4, results are tested for sensitivity to variation to cohort baseline settings and parameters. Section 5.5 summarises the main findings, with section 5.6 concluding the chapter.

5.2 Methods

The decision problem, that is, whether to allocate the resources required to implement LLGA programme, was evaluated from a health care sector perspective. A lifetime time horizon was selected to ensure that all relevant costs and benefits were taken into account. Incremental QALYs, LYs and costs were estimated using the decision analytic model developed in chapter 4. Further details on the analytical methods used for this base-case analysis are reported below.

Methods consistent with public health evaluation guidelines (National Institute for Health and Care Excellence, 2012) will instead be applied in the next

chapter, where a public health perspective on costs and outcomes is explored.

5.2.1 *Choice of framework*

Methods of EE have been aligned with previous similar studies (Cavill, 2011, Frew et al., 2014, Montes et al., 2012, Munro et al., 2004) to ensure a degree of comparability across EEs. In particular, a cost-utility analysis was conducted to assess whether LLGA was cost-effective, relative to a no LLGA intervention scenario, from a healthcare perspective over a lifetime horizon.

5.2.2 *Economic model overview*

Assessment of the value for money of LLGA was based on the sample of programme participants. The simulation thus started with a cohort of 51,874 healthy adults aged 39 years old and 62.4% female, with the model running until cohort members reached age 100 years. The baseline distribution of PA categories by IMD level is summarised below in Table 5.1:

Table 5.1 Frequency distribution of physical activity by IMD status

N=51,874	INA	INS	MOD	ACT
IMD non-deprived n=41,737 (80.5%)	28.1%	37.6%	21.7%	12.6%
IMD deprived n=10,137 (19.5%)	32.8%	34.9%	20.1%	12.2%

INA=inactive, INS=insufficiently active, MOD=moderately active, ACT=active

A usual practice scenario was chosen as the comparator. A discount rate of 1.5% for costs and outcomes was applied following relevant recommendations (National Institute for Health and Care Excellence, 2012).

5.2.3 Input parameters

Effectiveness parameters

Transition probabilities between the four PA states (i.e. effectiveness parameters) calculated in chapter 3 using the card swipe measure (see appendix G) were used to populate the decision-analytic model. As described in chapter 3, a last observation carried forward approach was used. Zero values were assigned to the change in PA category by service non-users (n=28,393), enabling the inclusion of the whole sample in the analysis (n=51,874). In other words, participants' missing follow-up values were replaced by the participant's baseline PA level. This assumes that LLGA participants could improve their baseline PA level only through regular participation to the free exercise sessions. This represents an unprincipled approach to dealing with missing data. In addition, this is likely to be conservative, as LLGA participants could have been prompted to exercise outside the free sessions (e.g. jogging). However, given a lack of data and relevant evidence on which to base such an assumption, a last observation carried forward approach was instead taken. This allows the risk of overestimation of the intervention effect to be limited, which could have otherwise arisen if changes in PA levels observed within the subgroup of service users were generalised to service non-users.

Intervention costs

Appendix I includes the financial audit reports provided by LLGA administrators which include the cost breakdown by project function/component. These reports were used to justify the cost of LLGA to

the City Council. In line with the approach currently adopted to inform

reimbursement decisions by the NHS (Department of Health, 2012),

the budget expenditure was assumed to represent the opportunity cost of

implementing the intervention, under a constrained budget with a current

£20,000 - £30,000 willingness-to-pay threshold range (Claxton et al., 2015).

Based on the information provided by the LLGA administrator in terms of

attributable costs, the intervention cost was estimated considering the cost

items listed in the financial reports that were related to the delivery and

promotion of the programme. Specifically:

- 80% of staffing (£755,841.46*0.8 = £604,673.17), which included

  managerial, administrative and technical personnel;

- 80% of marketing (£85,800*0.8 = £68,639.78), which included lunch

  campaign (e.g. leaflets, radio ads and billboards) and microsite;

- 100% loss of income £849,743, which represented the total estimated

  overhead costs from implementing the programme.

The total cost of LLGA calculated over the programme duration (39 months)

was therefore £ 1,525,055.95 (£604,673.17+£68,639.78+£849,743). The unit

programme annual cost was thus simply calculated by dividing this cost by

the number of participants and number of years of programme duration

(£1,523,055.95 / 51,874 / 3.25 = £9.03). Through this costing approach,

however, the perspective taken was that of the funder (i.e. Public Health

England and Sport England) rather than that of the body administering the

intervention (see next chapter, section 6.2.2.2). Furthermore, simply

accounting for the financial costs occurred within the programme duration made the cost estimate likely not to be a reliable measure of longer-term costs due to the intervention.

This financial costing approach, however, failed to capture differences in timing related to when costs of certain inputs incur and the rate at which capita items are used (Walker and Kumaranayake, 2002). Issues surrounding the handling of cost data have been discussed in the literature (Malehi et al. 2015). In particular, uncertainties relate to annualised costs and how to account for capital costs which typically occur at the beginning of an intervention, but the services from them could last several years. Considering the LLGA case study, it was assumed that recurrent costs (e.g. staff and leisure centre maintenance) would be similar each year and a constant rate of depreciation of capital items (e.g. the programme website). Furthermore, the intervention cost estimate was based on an aggregate measure of resource use and potential heterogeneity between leisure centres and service providers were not taken into account. These factors limit the validity and generalisability of this measure to other time periods (especially lifetime) and to other contexts. Nonetheless, this choice of method was dictated by the data available and it represents a common approach (Wolfensletter and Wenig 2011) which aligns with the analysis conducted by Frew et al. (2012).

5.2.4 Assumptions

Key assumptions were made in regard to the behaviour change measure used, the "natural course" of PA, the decay of effectiveness over time and mechanism to generate health improvement.

The main difference with previous EEs is the measure of behaviour change used. Previous models predominantly used self-reported levels of PA as outcome measures. Furthermore, the systematic review of current economic models presented in chapter 2 revealed that the vast majority of EEs based estimations of programme effectiveness on complete case analysis approaches. In other words, they assumed that no selection bias had occurred and generalised the observed results to the remaining of the population.

As mentioned above, in this base-case analysis, a more conservative last observation carried forward method was applied, as a form of intention-to-treat analysis. Participants could improve PA category only through a sustained rate of service use. In other words, it was assumed that the LLGA programme could not have affected PA behaviours otherwise (e.g. prompting individuals to exercise outside LLGA). However, by using this measure, it was assumed that attending LLGA sessions was additive to the other PA behaviours (e.g. active commuting), and that one LLGA session corresponded to at least 30 minutes of at least moderate PA (i.e. an active day).

In line with previous evaluations of universal programmes (Frew et al., 2014, Montes et al., 2012), a parallel trend assumption was made. In other words, that baseline PA levels did not change for participants that were not exposed to the intervention. Moreover, a no decay of intervention effect was also assumed, with benefits of improved PA being assumed to be immediate (i.e. no time lags between change in PA and health benefit) and sustained over the whole time horizon.

5.2.5 Sensitivity analysis

Parameter uncertainty relates to the accuracy and precision of the input data. This type of uncertainty was characterised using multiple methods (Briggs, 2006). Deterministically, scenario analyses, one-way and multi-way sensitivity analysis, and probabilistically using probabilistic sensitivity analysis (PSA). PSA captures any sampling uncertainty represented by uncertain distributions assigned to parameters.

Scenarios were chosen according to what was deemed important from a health policy perspective. In particular, the simulations were repeated for a younger (start age at 16 years old) and an older cohort (start age 65 years old), aligning the cohort in terms proportions of PA levels and socio-economic deprivation groups. Deterministic sensitivity analysis ranges were defined for parameters including: effect of the programme, intervention costs, disease risks and discount rate which were tested for 20 to 30% variations, in line with previous similar studies (Frew et al., 2012; Goyder et al. 2014).

A Monte Carlo simulation was used to propagate this uncertainty through the model and allow model parameters to vary simultaneously. Multiple iterations

are used to represent the full distribution of uncertain parameters. A thousand samples were simulated to assess the likelihood of the intervention to be the optimal alternative. Table J.1 (see appendix J) includes details on the distributional forms chosen for each set of parameters, together with the methods of estimation and moments values.

Under a standard assumption of joint multivariate normality, Cholesky decomposition of the variance-covariance matrix was used to capture correlation between regression coefficients. No uncertainty was assigned for the risks of mortality as these estimates are based on very large data sets (i.e. national-level registers). As mentioned in the previous section, modelling and methodological uncertainties are further explored in Chapter 6.

5.2.6 Threshold analysis

As discussed in Chapter 3, the effectiveness derived from the analysis of LLGA data were deemed likely to have been subjected to bias, due to lack of appropriate study design. However, decision-makers may be interested in knowing what would be the minimum level of effectiveness required for a programme like LLGA to be cost-effective. To this purpose, a threshold analysis was performed. Given the programme objectives, minimum effectiveness was calculated in terms of proportion and number of inactive residents needed to improve PA for the programme to be cost neutral (i.e. INMB equals zero). These estimates were calculated as the proportion of inactive adults, as well as by IMD subgroup (non-deprived and deprived).

The proportion of people transitioning between the inactive to the next higher PA state was varied progressively from zero to 100%. If a 100% transition to such PA level was not sufficient, the process was repeated for the next higher PA level (i.e. from inactive to moderately active) until break-even was reached. Three thresholds were considered: NICE's current upper bound (£30,000), lower bound of willingness to pay (£20,000) and opportunity cost per QALY faced by the NHS (£12,936) estimated by (Claxton et al., 2015) . Loss of monetary benefit was calculated if no break-even could be reached.

5.3   Deterministic results

5.3.1 *Cost-effectiveness outputs*

Table 5.2 shows absolute and incremental effects, costs and outcomes associated with a LLGA and a no-LLGA scenario, estimated over a lifetime time horizon.

Table 5.2 Per-participant cost-effectiveness outputs

N=51,874	LLGA	no-LLGA	Difference
LYs	38.0879	38.0876	0.0003
QALYs	25.9078	25.9054	0.0024
Costs	£158,494	£158,486	£8
Incremental cost per QALY gained	£3,239		
Cost-savings to the NHS*	£22		
Incremental Net Monetary Benefit**	£40		

* calculated as the difference in disease treatment and management costs between LLGA and no-LLGA intervention options; **calculated as the difference between value of a QALY gained (λ=£20,000) and incremental costs of LLGA (cost savings to the NHS – programme cost)

LLGA was found to be cost-effective, with positive incremental costs and positive QALY gains relative to a no-LLGA scenario. These results were based, however, on only small mean differences in PA. Only a small proportion of the cohort (529 of 51,874) improved their PA category. Implementation of LLGA was also associated with lower health care costs, relative to a no intervention option. These lifetime cost savings to the NHS were generated from the number of disease case averted (n=239 over 51,874 participants, see breakdown by disease in next chapter Table 6.4) and consequent lower use of health care resources for disease treatment and management. Table 5.2 also shows a small difference in terms of life expectancy projected between the two intervention options of 0.0003 life years (LLGA 38.0879 versus no-LLGA 38.0876). This was equivalent to extra 2.6 hours of life expectancy gained per participant (0.0003 x 365 x 24), on average. This limited differential effect on survival is driven by the assumption that there is no independent effect of physical activity level on mortality. Instead mortality changes can only be observed through disease diagnosis associated with PA levels.

If a QALY is valued at £ 20,000, for the entire cohort a total of around £ 2 million was estimated in terms of INMB over a lifetime, with £1.1 million cost savings to the NHS for disease treatment and management. A total of 239 disease cases and 17 deaths were projected to be averted over a lifetime (71 years). LLGA was associated with an improvement of one PA category by 176 inactive participants, at an incremental cost of £8,665 per inactive participant.

5.3.2 Threshold analysis

Table 5.3 below shows the number and proportions of inactive adults (i.e. zero active days) needed to improve their PA by one category (i.e. to become insufficiently active) for the programme to be cost-neutral, under base-case assumptions.

Table 5.3 Threshold analysis results

NHS threshold	Inactive (n=15,050)		Inactive non-deprived (77.9%, n=11,726)		Inactive deprived (22.1%, n=3,324)	
	Prop.	Number	Prop.	Number	Prop.	Number
£12,936	0.45%	68	0.56%	66	2.31%	77
£20,000	0.31%	47	0.39%	46	1.55%	50
£30,000	0.21%	31	0.27%	30	1.09%	36

The number-needed-to-treat for the programme cost to be counterbalanced by health benefits ranged from 30 inactive adults from non-deprived areas (0.26% of this subgroup) with a QALY valued £30,000, to 77 inactive adults from deprived areas (20.7% of this subgroup) with a QALY valued at £12,936.

Despite carrying an overall higher risk of disease, and consequently a higher potential for cost-savings to the NHS, a higher number of adults from deprived areas was needed compared to those from non-deprived areas. This is due to different utility values attributed to PA states between non-deprived and deprived (see appendix J), which award more the former group for equal changes in PA categories. For example, a transition from an inactive to an insufficiently active state corresponds to a utility gain equal to

0.0509 for the non-deprived group, while, for the same transition among the deprived, the gain is 0.0446 (i.e.13% lower).

From a policy perspective, an equal utility value for the deprived and non-deprived group meant that socio-economic differences do not matter at the lower bound of the physical activity spectrum. On the other hand, improvements in PA levels by the non-deprived group would generate more additional utility than the socio-economic deprived, making the task of reducing the inequality gap even harder from a public health standpoint.

5.4   Sensitivity analysis

5.4.1   Deterministic sensitivity

Figure 5.1 below shows the impact of deterministic sensitivity analysis relating to the characteristics of the cohort and values for input parameters. Also shown is results from a multi-way sensitivity analysis. The INMBs estimated for the base case (in £1,000; £2,791 at the top of the chart) are shown alongside the sensitivity analyses.

None of the alternative scenarios generated negative INMBs. LLGA was found to be generally robust to variations to both cohort characteristics and parameters, under base case assumptions. With regard to the former, an approximately £3 million difference in projected monetary benefits was estimated favouring a cohort of young adults (16 years old) over an older

cohort (65 years old). While the baseline difference in the proportion of IMD

levels (80.5% non-deprived vs 19.5% deprived) seemed to play a minor role,

around a fifth (27.6%) of the benefits estimated for the base case was

attributed to the baseline difference in distribution of PA categories between

IMD levels.

Figure 5.1 One way and multi-way sensitivity analysis
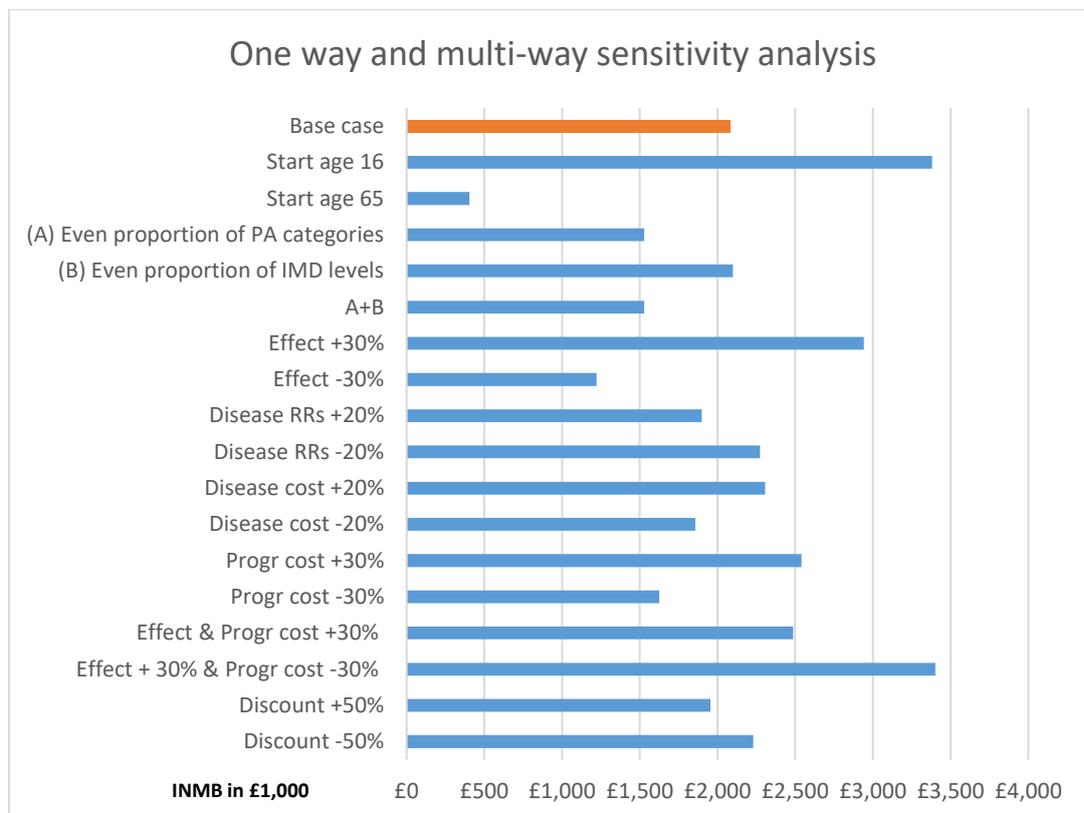
In terms of model inputs, effectiveness and programme costs appeared

major drivers of cost-effectiveness results. A 30% negative difference in the

proportion of inactive participants improving PA category would cancel out

the whole cost-savings to the NHS estimated for the base case (£1.1 million).

When this difference in the proportion of inactive participants was combined

with a 30% reduction in programme costs, an INMB of £3.4 million was

estimated.

5.4.2 Probabilistic sensitivity

Figure 5.2 below shows one thousand model iterations of the cost and QALY joint density plotted on a cost-effectiveness plane, comparing LLGA intervention to a no-LLGA scenario (set as the origin), for a lifetime horizon. The simulation produced a fairly dispersed cloud of points which fell mostly on the East quadrants, indicating that LLGA was highly likely to generate QALY gains. This level of dispersion indicated a relatively high level of uncertainty around the expected ICER values. Looking at the distribution of cost and QALY pairs, the majority fell below the lower bound willingness-to-pay threshold, indicating that there was a high probability of LLGA being the optimal alternative.

Figure 5.2 Cost-effectiveness plane
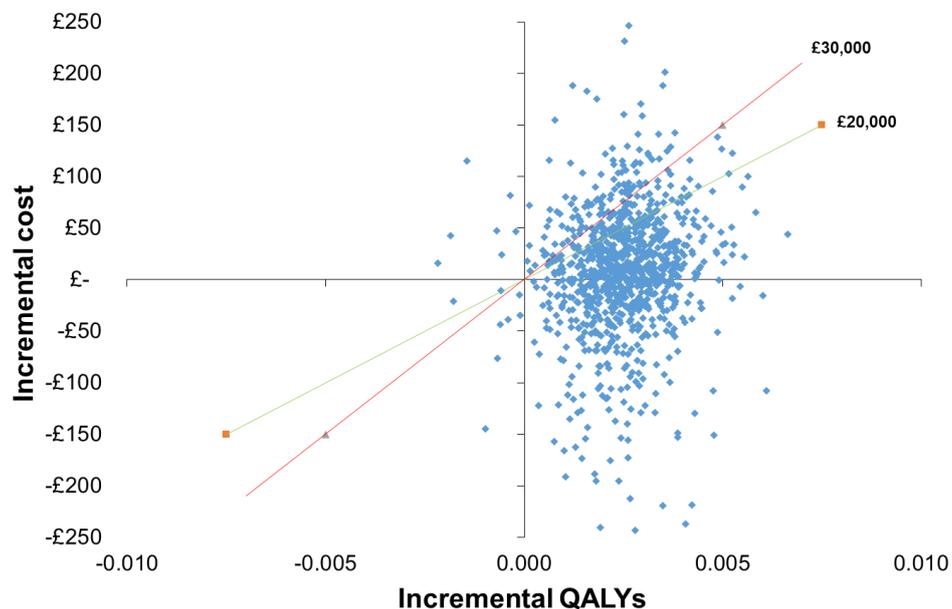
Figure 5.3 shows below the probability of LLGA being cost-effective, across a range of willingness-to-pay thresholds.

The cost-effectiveness acceptability curve (CEAC) did not cut the y-axis at zero (i.e. 50%) indicating that part of the joint density involved cost-savings. Only part of the density involved QALY gains, as apparent from a CEAC converging relatively slowly to probability of 1 (Fenwick et al., 2004). Reflecting what was displayed in Figure 5.2, a relatively high (80%) probability of LLGA being the optimal strategy was found when considering a £20,000 threshold.

Figure 5.3 Cost-effectiveness acceptability curve
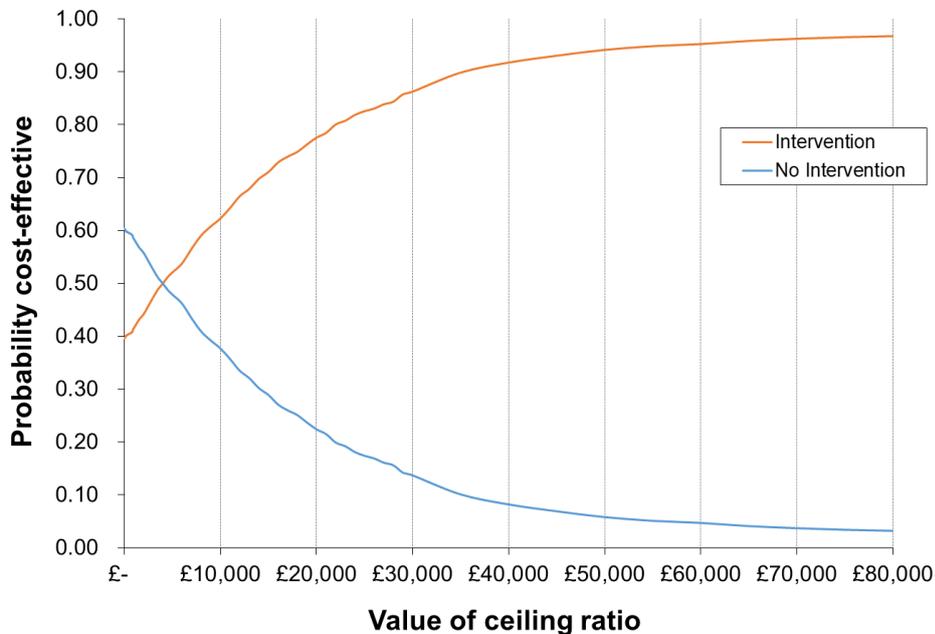
5.5 Main findings

Cost-effectiveness results suggested that LLGA programme was likely to be cost-effective, compared to a no-intervention scenario, under base case assumptions. This was despite having used a more conservative measure of change in PA behaviour, compared to previous EEs. Results from a threshold analysis confirmed that a relatively small, but sustained

improvement in PA at a population level could generate substantial net benefits in the long-term. Results showed high sensitivity to variation to programme effectiveness and costs, while being robust to probabilistic uncertainty.

5.6 Chapter summary

This chapter presents an EE of the LLGA programme. This analysis was conducted with the purpose of obtaining base-case estimates of cost-effectiveness. In particular, key modelling and methodological assumptions have been aligned with those of previous studies to ensure a degree of comparability with their results. This was done to allow for a degree of generalisability of the scenario analysis results generated in the next chapter, where four sets of modelling and methodological assumptions are tested (as identified at the end of chapter 2). In particular, the next chapter provides quantitative evidence on the impact that variations to four sets of assumptions, which characterised previous similar studies (chapter 2), can have on cost-effectiveness, illustrating their implications for decision-making.

6   Chapter

Testing the assumptions that characterise the evaluation of universal programmes

6.1   Chapter outline

Moving the focus from the cost-effectiveness of the LLGA programme to the EE of universal programmes to promote healthy behaviours more broadly, this chapter illustrates the implications of four sets of modelling and methodological assumptions on cost-effectiveness. These are: mechanisms of second-stage survey non-response, behaviour change maintenance over time, range and combination of modelled diseases, and perspective for EE.

Before proceeding with the analysis, section 6.2 describes how the modelling assumptions are tested and the methods used to explore changing the perspective for EE. In respect to the latter, a HWB and a Local Authority perspectives are explored. Section 6.3 illustrates the implications of changes to the investigated assumptions. Section 6.4 summarises the main findings and 6.5 concludes the chapter summarising its main points.

6.2   Methods

The implications of variations to these assumptions is explored in the following sections by showing their impact on cost-effectiveness outputs obtained from the EE conducted in the previous chapter. NICE recommends the exploration of uncertainty to adequately inform decision-making (National

Institute for Health and Care Excellence, 2012). Specifically, separate analyses of a range of scenarios are recommended to be conducted (National Institute for Health and Care Excellence, 2013). Scenario analysis was used to explore these variations, with tables and graphic representations being used to describe differential outcomes, as appropriate.

In order to support the interpretation of results, a net benefit framework was used (Hoch et al., 2002). In particular, the incremental net monetary benefit (INMB) of implementing LLGA was calculated for each of the alternative assumption scenarios.

Given

INMB = Δ effectiveness*λ – Δ costs, with λ set at £20,000.

a positive INMB indicated that the LLGA is cost-effective compared with no intervention and viceversa.

6.2.1 Modelling assumptions

6.2.1.1 Mechanisms of second-stage survey non-response

Four alternative mechanisms were simulated, corresponding to the four adjustment methods detailed in chapter 3, using the estimates obtained from analysis of the survey data (i.e. survey measure).

These were: 1) last observation carried forward, under a pragmatic assumption that those adults for whom no second-stage ("follow-up") data were available did not change their baseline PA, despite being exposed to LLGA offer (i.e. signed up to it); 2) complete case analysis, under a MCAR assumption meaning that the change observed in survey respondents

represented that of the whole sample; 3) inverse probability weighing, under a MAR assumption meaning that the change observed in survey respondents was only influenced by, and therefore adjusted, for observed factors; 4) Heckman selection model, under a MNAR assumption meaning that the change observed in survey respondents was only influenced by, and therefore adjusted, for unobserved factors.

6.2.1.2   Range and combination of modelled diseases

As identified in the review of previous models (chapter 2), high variability in the range and combination of modelled diseases was found across studies. The decision-analytic model developed here (chapter 4), which was used for EE of the LLGA programme, included most of them, with frailty syndrome being also included for the first time.

To illustrate the impact of this method choice, diseases were grouped by type of condition: metabolic (type II diabetes), cardiovascular (coronary heart disease and stroke), genetic mutation (colorectal and breast cancer), mental (depression) and geriatric (frailty) and added incrementally to the decision model.

6.2.1.3   Behaviour change maintenance over time

For the base case analysis of LLGA programme (chapter 5), in line with previous evaluations, a no decay of effect assumption was made. In other words, the change in PA category observed within the assessment period, and the derived health benefits, were assumed to remain constant at 100% over the whole time horizon. As mentioned in the literature review chapter,

this represents a fairly unrealistic assumption, which is usually dictated by a lack of relevant data.

In an effort to make the best use of the available data, two alternative scenarios, namely, immediate and exponential rebound were simulated. With an immediate rebound trajectory, no residual effect was assumed beyond the first 6 months since the programme started (i.e. beyond the assessment period). As for the latter, an exponential trajectory was derived by fitting a survival model to the card swipe data (chapter 3.4.2.2, "service use drop-off pattern"), following the procedure described in chapter 4.2.3.2 "residual intervention effect"). Under this scenario, it was assumed that the change in *overall* PA estimated in the cohort within the first 6 months would gradually return to zero, following a LLGA session attendance drop-off trajectory (see chapter 3.4.2.1).

6.2.2 *Perspective for economic evaluation*

For the base-case CEA of LLGA, choice of analytical methods followed that of a health care sector perspective. This entailed the selection of a lifetime time horizon, a focus on health gain maximisation and a budget allocation approach to costing the intervention. However, universal programmes in general, and LLGA specifically, do not usually involve only decision-makers from the health care sector. As mentioned in the Introduction chapter, under the new NHS structure (UK Government, 2012), since April 2013, Clinical Commissioning Groups are responsible and influence commissioning

decisions related to public health as part of local HWBs. HWBs are formal committees of the Local Authorities (e.g. local City Councils).

6.2.2.1   Health and Wellbeing Board perspective

Commissioning cycles of Local Authorities are short, with decisions covering financial frameworks from 3 to 5 years, being also dependent on funding cycles (Local Partnerships, 2014). Moreover, as public health agencies, HWBs have a dual objective of improving population health and reducing health inequities (The King's Fund, 2016). In addition, HWBs may be particularly interested in knowing whether a programme such that of LLGA is good value for money in the short term, for instance, to justify further funding. Therefore, a short time horizon was considered. Given LLGA budget life cycle and local authority's planning horizon, a 39-month time horizon was chosen, matching LLGA programme duration.

6.2.2.1.1   Health inequality impact

The impact of LLGA on baseline health inequality was assessed, following a sub-group CEA, in line with stated objectives by the local HWB. According to the Leeds City Council Executive Board report (see appendix K), LLGA offer was corroborated by the significant health and life expectancy inequalities which exists within Leeds, despite being the 7th most active Local Authority in England out of 326. Indeed, a difference of around 10 years in survival terms, and 20 years when considering healthy life expectancy have been estimated between the top and bottom decile of the population (Public Health England, 2015). This justified the choice of undertaking a health inequality analysis, in

line with the decision-makers' goal of assessing the distributional impact of LLGA.

In recent years, a number of approaches have been developed to this purpose (Cookson et al., 2017). Extended CEA methods, developed by Disease Control Priorities (Verguet et al., 2016), have been applied to the study of distributional impact of policies of health benefits and financial risk protection benefits in low and middle income countries, where the prevention of medical impoverishment due to medical costs is a major concern.

Other forms of distributional impact analysis have been conducted outside the context of CEA. For instance, benefit-incidence analysis to look at the relative benefits of public health care expenditure (Bowser et al., 2019) or examine the changes in risk factors and treatment utilization (Bajekal et al., 2012). However, these two approaches cannot provide decision-makers with relevant information as to which of the possible alternatives maximise the objective function, which is often unknown in public health.

Another framework developed by University of York is distributional cost-effectiveness analysis (DCEA), which was chosen for this analysis (Asaria et al., 2016b). Unlike the other methods, this extra-welfarist based approach considers not only the distributional impact of an intervention in health-related outcomes (i.e. gross health benefit), but it also reflects health equity implications of the distribution of health opportunity costs. This distribution is dependent on how the intervention is funded and may not be even across

population sub-groups. DCEA also allows possible trade-offs between health maximisation and equity objectives to be explored. These can be presented graphically via the health equity impact plane, or in a more sophisticated fashion, through aggregate measures of net health equity impact (Cookson et al., 2017).

6.2.2.1.2 Inequality measures

The distributional impact of LLGA on LYs and QALYs (i.e. average health gains per unit) was computed by IMD subgroup. Gross and net health inequality impact on QALYs were calculated as average (per unit) differences between the two subgroups, without and accounting for their health opportunity costs, respectively. The decision-analytic model accounts for size and group distribution, therefore, the health benefits can be directly used to calculate the impact on quality-of-life adjusted life expectancy (QALE).

QALE can be defined as the number of years an individual is expected to live in full life. This measure was selected due to its relevance in the context of LLGA programme, where the interest was to capture the difference projected between a LLGA and a no-LLGA options in terms of survival and quality of life by the cohort. QALE was calculated following the approach used by (Love-Koh et al., 2015). That is, it was assumed that same IMD-group members experienced the same average health-related quality of life for the period before the intervention started.

To calculate health opportunity costs, additional costs were converted using an opportunity cost of health of £20,000. Estimates of marginal changes in health expenditure by IMD subgroup were sourced from the analysis performed by James Love-Koh (2017). Absolute and relatives difference in QALE between the two subgroups were used as the inequality measures to describe the net health inequality impact of the intervention.

6.2.2.2 Local Authority perspective

Considering an alternative decision maker's perspective, the Local Authority requires the consideration of two concurrent responsibilities: promoting public health and service provider. While having an interest in achieving goals of population health as a component of local HWB, the Local Authority was also in charge of administering and hosting the programme within leisure centres that were managed by them. As a result, if a Local Authority perspective is assumed, the economic cost associated with implementing the intervention may differ from the estimated budget expenditure. In particular, the latter may not overlap with the opportunity cost faced by the Local Authority, in the case this public body was the only agency in charge of the decision.

In addition, as with HWBs, Local Authorities may be particularly interested in in the short term implications of implementing a programme such that of LLGA. To address such perspective, conditional on the information available, a scenario when the Local Authority does not have external financial support or decision influence was simulated.

6.2.2.2.1   Opportunity cost of the intervention

In the previous chapter, the economic evaluation was conducted from a healthcare sector perspective. The opportunity cost of implementing LLGA was assumed to be equal to the budget spent by the Local Authority to provide the free sessions and manage the programme (£1,525,000). In turn, the Sport and Active Lifestyles department of the Local Authority costed the intervention using the allocated budget following internal accounting guidelines (details not disclosed), similarly to that followed for the programme evaluated by Frew et al. (2014).

The only source of information relating to the resources used for implementing LLGA was represented by the financial audit reports shown appendix I. This limited the ability to estimate the opportunity cost of the intervention from a Local Authority perspective accurately. Based on these audit reports and discussion with the programme administrator, the costing approach adopted resembled that of full absorption costing. A full absorption approach includes not only the costs of material and labour needed for providing the service, but also all overhead costs. Such approach is recommended as the preferred method in the NHS costing manual (Department of Health, 2012).

Worthy of note in these audit reports is a cost item labelled as "loss of income", which accounted for a large proportion of the total cost (around 70%). This entry represented an estimate of the cost for provision of the LLGA sessions. According to the programme managers, this cost was

calculated by multiplying an average reference cost per session (£1.68) by the number of LLGA sessions *accessed* by participants throughout the programme. While no detailed information was provided in terms of how this reference cost was calculated by the Local Authority, it appeared to represent the average financial cost per routine session offered to the leisure centre members (i.e. outside LLGA programme).

Going back to its definition, the economic (or opportunity) cost of a decision is the value forgone as the result of opting for the best alternative option (Alastair M. Gray, 2011). Within the studied context, this cost encompassed two aspects: the value of additional resources needed for the intervention, and the net benefit lost from the next best alternative use of labour and capital involved. Given the nature of the intervention (i.e. promotion and offer of off-peak exercise sessions held in addition to currently scheduled sessions) and scope of this analysis, it was assumed that a no-intervention option was the only possible alternative. Therefore, no relevant benefits would be lost from implementing the intervention. As a result, the opportunity cost was equivalent to the additional resources used multiplied by the respective unit costs.

To calculate this opportunity cost, a series of steps were followed (Alastair M. Gray, 2011). First, identification. Sources of resource use were identified considering both the Local Authority's structure, relevant functions and potentially impacted activities. These are summarised in Table 6.1 below.

Table 6.1 Sources of resource use for intervention

		Function		
		Local Authority organisation	*Programme promotion*	*Programme delivery*
Activity	**Administration**	Internal audit, archive	Business administration	Business administration
	Communication	Internal reporting	Recruitment of volunteers, community engagement	Leisure centre management
	IT	Electronic data management	Programme website	Electronic cards and access data management
	Technical service	Office management	Mass and social media campaigns	PA professionals, project management and facility maintenance

Through this classification approach, resource use functions, rather than items, were defined as the primary drivers of cost. For example, costs for transportation were added to the identified activities/functions (e.g. for attending meetings or events), rather than being computed and categorised as accessory costs.

Table 6.2 Estimated cost of resources used for intervention

Source of cost	*Quantification method / cost driver*	*Unit cost (n=51,874)*	*Notes*
Programme website	Local Authority reports	£0.16	Total expenditure for website design and management £17,000: assumption 50%
Media campaigns	Local Authority reports	£0.66	Total expenditure £ 68.799,72: assumption 50%
Project management	Local Authority reports	£3.25	Total expenditure for project managers £168,654.50
Physical activity professionals	Scheduled programme sessions	£4.92	142 hourly sessions, 170 weeks, £22000 annual salary= £10.58 hourly wage (National Careers Service, 2018)

With regard to measurement, given the lack of more granular data, a top-down approach was used to estimate the per-person programme cost. A top-

down approach consists in allocating resource use to macro functions of the programme, rather than individual tasks (i.e. bottom-up approach, Olsson, 2011). As shown in table 6.2, four sources of cost were identified and measured: programme website, media campaigns, project management and PA professionals.

Considering that the programme website and the media campaigns were intended to promote other initiatives (i.e. community programme), and based on what has been reported by the programme administrator, an assumption of 50% of the related cost borne for the intervention was made. Project management staffing costs were based on the respective proportion of the budget outlay, while market pricing was used to value provision of the PA sessions.

Through a marginal costing approach, a total incremental cost of £8.99 per participant was estimated (annual £2.77, £8.99/3.25), with the related activities assumed to be all outsourced. In other words, it was assumed that the implementation of LLGA would not alter significantly (in economic terms) the structure of the hosting organisation (e.g. top management, infrastructures) beyond what was taken into account in Table 6.2, therefore, overhead costs were not included.

6.3   Results

6.3.1 Mechanisms of second-stage survey non-response

Table 6.3 below shows the INMB values associated with four alternative assumptions regarding the causes of missingness in the survey measure, analysing the sample of LLGA participants (N=51,874) over a lifetime. The first row reports the base-case results estimated in chapter 5 using the card swipe measure for comparison.

Table 6.3 Impact of assumptions about survey non-response on INMB

	Assumption	**Adjustment method**	**INMB**
Base-case results (chapter 5)	Service non-users did not change PA category, if not though LLGA	Last observation carried forward	£ 2,075
	Survey non-respondents did not change PA category	Last observation carried forward	£ 368
	MCAR, no selection bias occurred	Complete case analysis	£ 508,557
	MAR, selection bias from and adjusted for observed factors	Inverse probability weighing	£ 453,008
	MNAR, selection bias from and adjusted for unobserved factors	Heckman selection	£ 552,513

Notes: MAR=missing at random, MCAR=missing completely at random, MNAR=missing not at random; values in £1,000

Compared to the results obtained using card swipe data (£ 2,075 million, based on change in PA by 529 service users), the change in PA category reported by survey respondents (n=547) led to a positive, but significantly smaller INMB (£ 368,378). This was due to fact that with the survey measure, unlike with the card swipe measure, participants could also self-report themselves at a lower level of PA after LLGA, therefore generating losses in QALYs. Given the structure of the card swipe measure instead, participants could not remain at the same PA level or improve it from baseline.

As expected from the difference in magnitude observed between the average estimates of effectiveness parameters computed in chapter 3, a very large difference in projected INMBs was found between the most conservative approach (last observation carried forward) and the other three approaches. Departures from a MCAR assumption generated divergent results, with a MAR approach correcting the MCAR results downwards by 11% and the MNAR model correcting them upwards by 8.6%. While the differences estimated between the three formal approaches were relatively small in terms of average changes in PA behaviour (see chapter 3.6.1), these corresponded to large differences in projected INMBs, in the order of £50 - £100 million over a lifetime.

6.3.2 *Range and combination of modelled diseases*

Figure 6.1 below shows the trajectories of disease cases averted over a lifetime for each of the seven chronic conditions modelled. The area between start age (39 years) and programme end (vertical red dotted line) has been zoomed for a clearer representation of the short term projections. The dark red dotted line represents the trajectory of disease cases averted if frailty was the only condition modelled.

Figure 6.1 Trajectories of disease cases averted per chronic condition
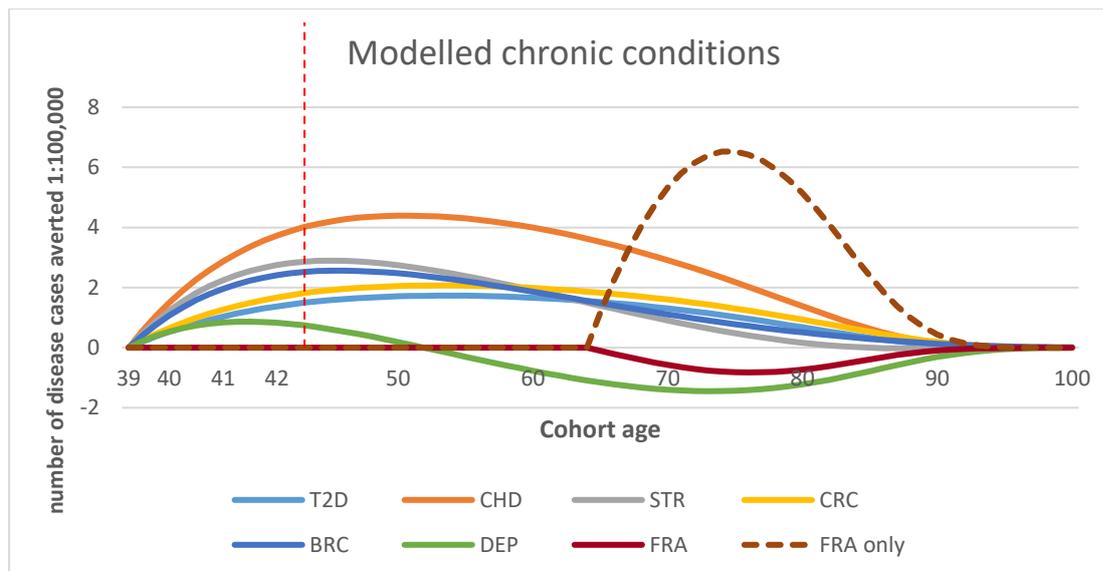
Note: T2D=Type 2 Diabetes; CHD=Coronary Heart Disease; STR=Stroke; CRC=Colorectal Cancer; BRC=Breast Cancer; DEP=Depression; FRA=Frailty syndrome.

Over the short term, all the relevant 6 diseases contribute positively to the accumulation of health benefits. This remains true until the cohort reaches 52 years, when the number of cases of depression starts accumulating, instead of decreasing, compared to a no-intervention scenario. This is because of the greater preventive effect of the intervention on the other diseases. This leads to a higher number of (alive) healthy individuals to accumulate over time in raw numbers, compared to a no-intervention scenario, who can move to a depression state in a progressively greater number.

Frailty, which by default settings is allowed to start its action only at 65 years of cohort age, starts and concludes its trajectory below the x-axis, gradually converging to the x axis with the other diseases as the cohort reaches age 100 (when all remaining members are assumed to die).

Figure 6.1 also shows the dynamic interactions across the seven chronic conditions, which compete one another in their probability of occurring. The

magnitude of the competing risk introduced by modelling more than one disease at a time is a function of the number of diseases modelled, their differences in terms RRs between PA categories and risk sizes. This effect is apparent from this Figure, which compares the trajectory of frailty cases averted if this condition was modelled in combination with and without the others.

Table 6.4 below shows the relative changes in the number of cases averted for each disease, when including additional sets of diseases.

Table 6.4 Impact of range of diseases on cases averted and INMB (N=51,874)

Range of disease	Disease	Number of cases averted	total n. of disease cases averted	INMB in £ 1,000*	
(Base case approach)	T2D	36	239	£	2,082
	CHD	88			
	STR	45			
	CRC	44			
	BRC	46			
	DEP	-13			
	FRA	-7			
T2D only	T2D	102	102	£	1,186
T2D + CHD, STR	T2D	56	291	£	1,684
	CHD	144			
	STR	91			
T2D,CHD, STR + CRC, BRC	T2D	41	297	£	2,534
	CHD	100			
	STR	52			
	CRC	51			
	BRC	52			
T2D,CHD, STR, CRC, BRC + DEP	T2D	36	244	£	2,177
	CHD	88			
	STR	45			
	CRC	44			
	BRC	46			
	DEP	-13			
FRA only	FRA	54	54	£	1,420

As showed graphically in Figure 6.1, depression and frailty contribute negatively to the total number of disease cases averted. When modelling

171

T2D only, the number of cases averted reaches 102. This number almost triples (n=291) when adding the two cardiovascular conditions, the absolute risks of which are four to five times greater than T2D. When further adding the two cancers, which carry risk sizes smaller than the cardiovascular conditions, the total number increases by only 5 cases, with the previous three conditions being downsized in their marginal effects.

The inclusion of depression to this disease set (T2D+CHD+STR+CRC+BRC) results in a reduction of the marginal (by each condition) and total number of cases averted (n=244). As mentioned above, this is due to the fact that changes in PA category has a smaller effect on depression (RR for inactive=1.15), compared to the other conditions, which counterbalances the other greater preventive effects. As a result, while in the first period the change in distribution of PA categories generates an increase in the number of cases averted, the stronger preventive actions by the other diseases (due to higher RRs) leads progressively to a higher number of healthy cohort members, relative to the no-intervention arm.

In terms of INMB, the different combinations of conditions also result in differences in predicted monetary values. These values depend on the effects described above and on the differences in costs for disease treatment and management avoided. This is apparent from Table 6.3, which shows that if, for instance, only T2D is considered, 102 cases are averted generating £1,186,000 (average £11,627).

If instead only frailty syndrome is modelled, the 54 cases averted generate £ 1,420,000 (average £26,296), that is, more than double the value per case of the previous scenario. This difference is even more marked with the inclusion of the two cancers (from section 3 to 4), contributing to only 5 more cases averted, but additional £221,000 (i.e. average of £44,200 per disease case averted).

6.3.3 *Behaviour change maintenance over time*

Figure 6.2 shows the trajectories of disease cases averted, compared to no-intervention, by assumption on behaviour change maintenance over time: no decay of intervention effect (base case, orange line), an immediate rebound (no residual effect first after 6 months, grey line) and an exponential rebound trajectory ("programme attendance drop-off", obtained fitting the card swipe data, see Figure 3.1).

The three assumptions on the longer-term sustainability of the intervention effect lead to very large differences in number of cases averted (as difference in areas under the curves).

Figure 6.2 Assumptions on behaviour change maintenance over time
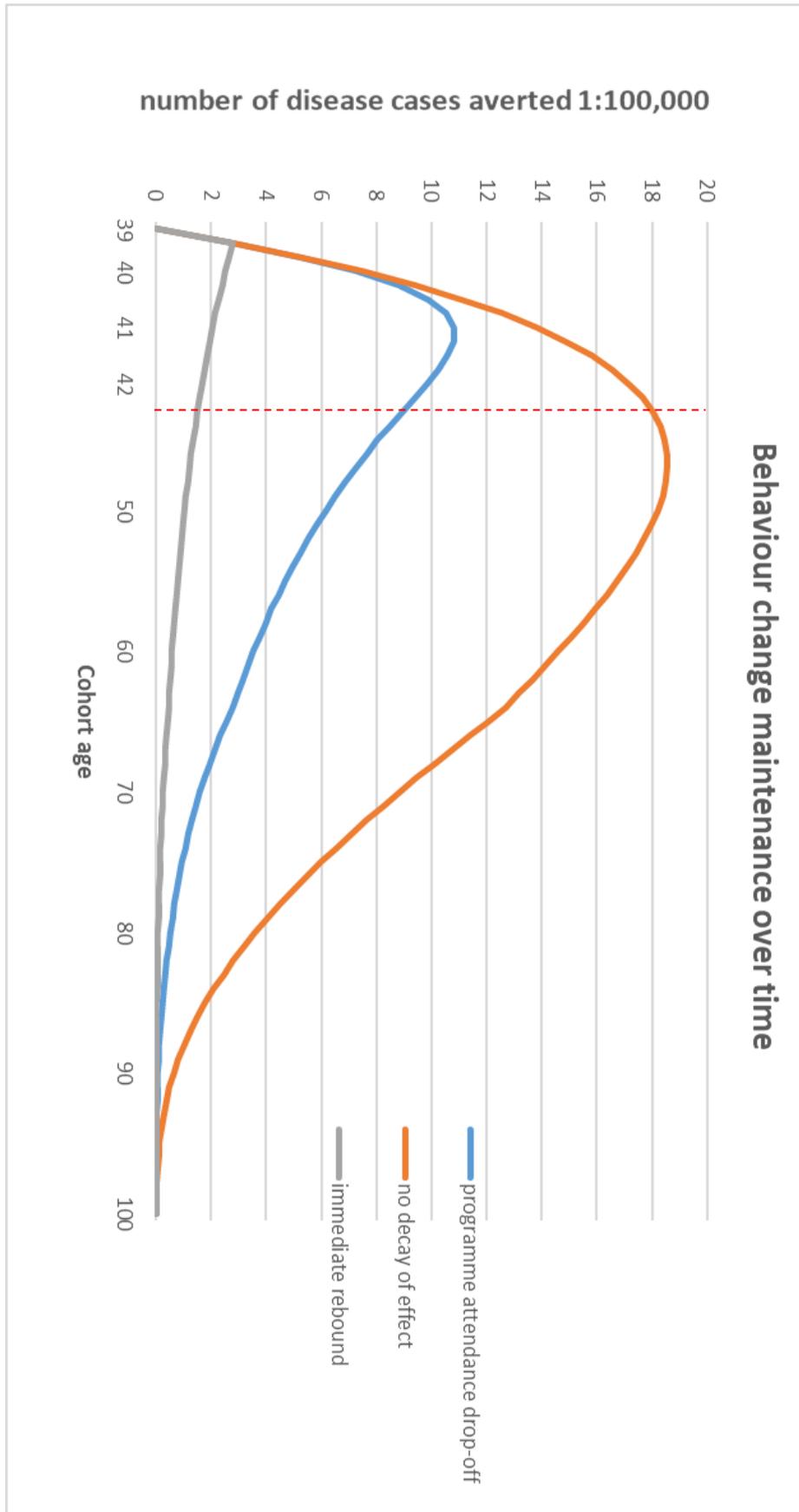
Under the most optimistic scenario (no decay of effect), the rate of disease cases averted keeps rising, reaching a peak after around 10 years from the start of the intervention, to then gradually converge towards zero. To a lesser degree, this rate follows a similar pattern under an assumption of a decay of behaviour change being equal to programme attendance drop-off, reaching its peak at around 2 years. Under the strictest assumption of no residual effect after the first 6 months, while the rate starts decreasing immediately after, a positive contribution to the number of disease cases averted is made, relative to a no-intervention scenario, until the cohort reaches 80 years.

Figure 6.3 Impact of behaviour change maintenance assumptions on INMB

Impact of behaviour change assumptions on INMB
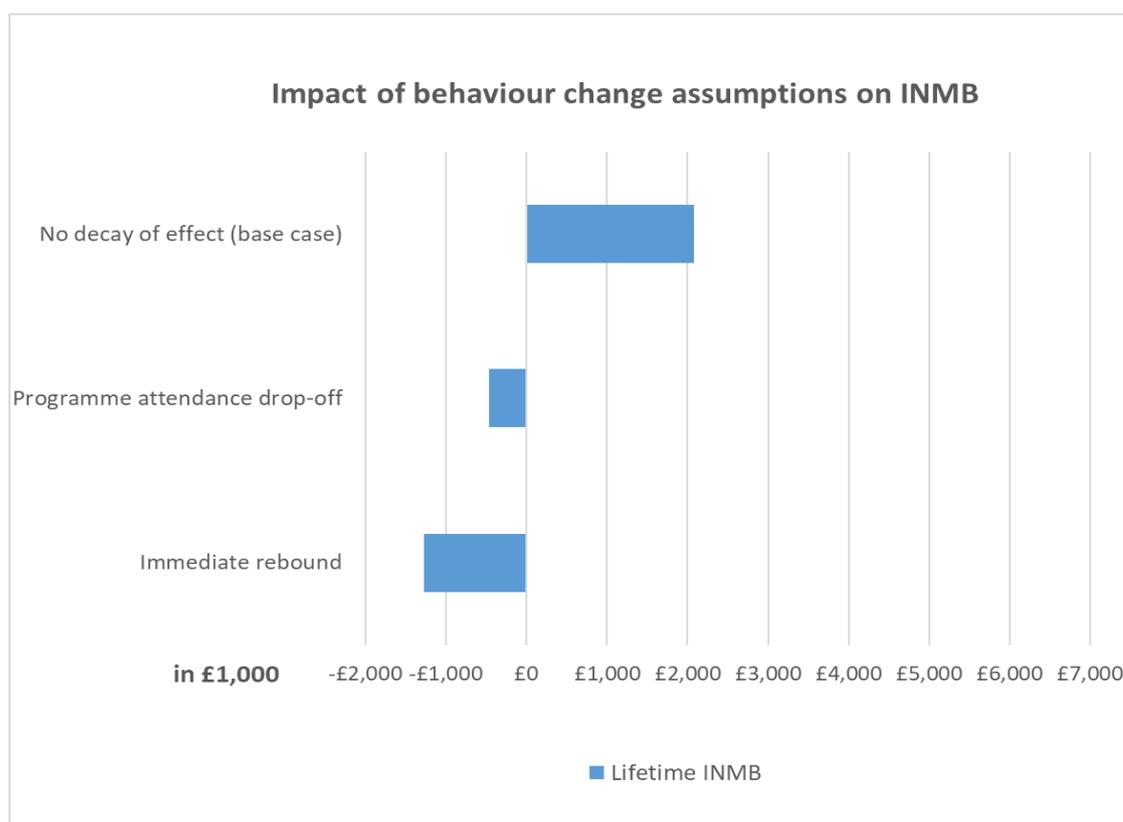
Figure 6.3 above illustrates the impact of these three assumptions on INMB. The differences in frequency distribution of PA categories and number of

disease case averted between these assumptions result in different utility gains associated with implementing the intervention.

Under the two alternative assumptions of immediate and exponential rebound, the intervention is projected to generate a negative INMB over a lifetime, with a difference of around £0.9 million between the two scenarios. This means that, unless the intervention effect is assumed to last beyond its duration (i.e. 39 months), the health benefits generated by the programme do not exceed the incremental costs needed for its implementation. However, under an exponential rebound trajectory assumption, the programme almost reaches a break-even point, with a negative INMB of £54,427 (result not showed, £1.05 per-participant cost), if a QALY is valued at £30,000.

6.3.4 Perspective for economic evaluation

6.3.4.1   Health and Wellbeing Board perspective

The next section illustrates the cost-effectiveness results estimated when the simulation is run for a short term time horizon of 39 months. The following section focuses on the impact of LLGA on baseline inequality in QALE between the two IMD groups (non-deprived and deprived), with results showed for both a lifetime and a 39-month time horizon. Analyses are based on the sample of LLGA participants (N=51,874).

Table 6.4 shows per-participant absolute and incremental costs and outcomes associated with LLGA and a no-intervention scenario, comparing a lifetime versus a 39 month time horizon.

Table 6.4 Cost-effectiveness outputs: lifetime versus 39-month time horizon

	Lifetime			39 months		
N=51,874	LLGA	no-LLGA	Difference	LLGA	no-LLGA	Difference
LYs	38.0879	38.0876	0.0003	3.2413	3.2413	>0.0001
QALYs	25.9078	25.9054	0.0024	2.9997	2.9995	0.0002
Costs	£158,494	£158,486	£8	£3,647	£3,619	£28
Incremental cost per QALY gained	£3,239			£115,230		
Cost-savings to the NHS*	£22			£1		
Incremental Net Monetary Benefit**	£40			-£23		

*calculated as the difference in disease treatment and management costs; **calculated as the difference between value of a QALY gained (threshold £20,000) and incremental costs of LLGA (cost savings to the NHS – programme cost)

Comparing with the (per-participant) outputs calculated over a lifetime, a very different picture of the impact and cost-effectiveness of the intervention is found. Over a 39 month time horizon, the intervention is associated with positive QALY gains, which represent around 8% of the QALY gains projected over a lifetime. LLGA is associated with a greater incremental costs (£28 per participant), and expected ICER of over £115,000, making this alternative cost-ineffective, at the current willingness-to-pay threshold.

This difference in economic efficiency is apparent from a negative INMB estimated at around £23 per participant (around £1.2 million for the whole sample of participants). This result starkly contrasts with that estimated for the whole sample over a lifetime in terms of projected INMBs, which was positive and estimated at around £2.1 million.

Large differences between the two time horizons also emerge in terms of

uncertainty around the mean ICER and probability of the intervention being

the optimal alternative.

Table 6.5 Cost-effectiveness plane: 39-month time horizon
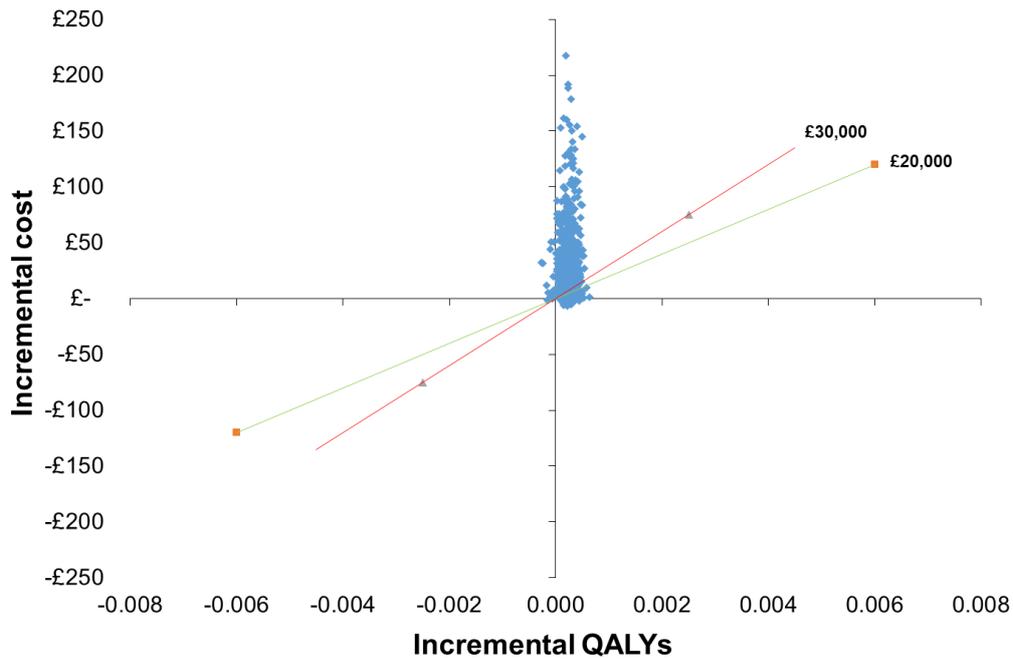
Unlike what found analysing a lifetime time horizon (see Figure 5.2), the

Monte Carlo simulation produced a compact cloud of points which fall almost

entirely on the North-East quadrant and above the thresholds. This

suggested that the probability of the intervention to be the optimal alternative

was low.

Looking at the joint probability distribution of incremental costs and QALYs

(see Figure 6.4 below), the cost effectiveness acceptability curve crosses the

y-axis just above the origin, indicating that cost-savings play a relatively

marginal role. Furthermore, LLGA becomes the optimal strategy only at willingness-to-pay threshold values higher that £81,000.

Figure 6.4 Cost-effectiveness acceptability curve: 39-month time horizon
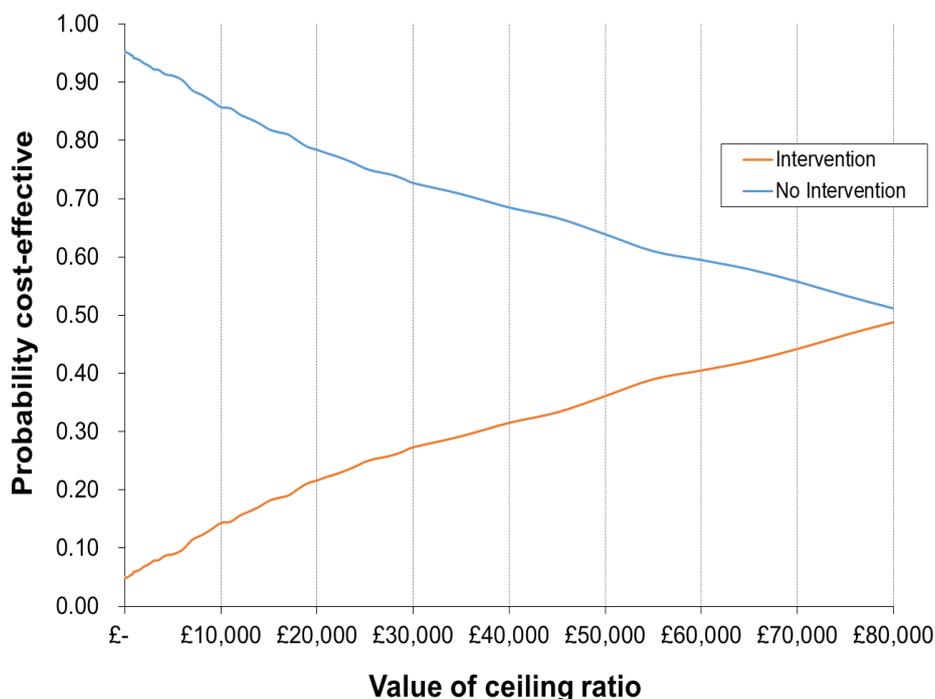
6.3.4.1.1   Health inequality impact

Before focussing on the impact of the intervention on baseline health inequality between the two IMD groups, Table 6.6 below summarises sub-group cost-effectiveness results estimated over a lifetime and a 39-month time horizon.

The intervention is found to benefit adults from deprived areas to a greater extent, at a higher cost, relative to those from non-deprived neighbourhoods. However, in absolute terms, the largest proportion of health benefits are accumulated by the non-deprived group who represent the majority in the

population (80.5%). Negative INMB estimates are found in the shorter term, making the intervention not cost-effective for either of the two subgroups. By contrast, lifetime ICERs show that the intervention would be a cost-effective alternative, with either of the two IMD groups.

Table 6.6 Sub-group cost-effectiveness outputs: lifetime versus 39-month time horizon

n=51,874	Lifetime		39 months	
	NON-DEPRIVED (n=41,737; 80.5%)	DEPRIVED (n=10,137; 19.5%)	NON-DEPRIVED (n=41,737; 80.5%)	DEPRIVED (n=10,137; 19.5%)
LYs	0.0003	0.0003	>0.00001	>0.00001
QALYs	0.0023	0.0026	0.0002	0.0003
Costs	£ 7	£ 11	£ 28	£ 28
Incremental cost per QALY gained	£ 2,952	£ 4,297	£ 118,766	£ 102,634
Cost-savings to the NHS*	£ 22	£ 18	£ 1	£ 1
Incremental Net Monetary Benefit**	£ 40	£ 41	- £ 24	- £ 23

*calculated as the difference in disease treatment and management costs; **calculated as the difference between value of a QALY gained (λ=£20,000) and incremental costs of LLGA (cost savings to the NHS – programme cost).

Implementing LLGA results in QALY gains over no intervention in both time horizons, favouring the deprived over the non-deprived group (i.e. positive gross health inequality impact). When accounting for sub-group health opportunity costs of implementing the intervention, two opposite results were found between the two considered time horizons. Table 6.7 below shows the distributional impact of LLGA, comparing the two considered time horizons.

Table 6.7 Distributional impact of LLGA programme

n=51,874	Lifetime		Budget life-cycle	
	NON-DEPRIVED (n=41,737; 80.5%)	DEPRIVED (n=10,137; 19.5%)	NON-DEPRIVED (n=41,737; 80.5%)	DEPRIVED (n=10,137; 19.5%)
Gross health inequality impact	0.00027		0.00004	
NON-DEPRIVED opportunity cost	0.00014			
DEPRIVED opportunity cost	0.00019			
Baseline QALE inequality gap	1.09054			
Net health inequality impact	0.00022		-0.00002	
Post-intervention QALE inequality gap	1.09033		1.09056	
Relative difference in QALE	0.020%		-0.002%	

Notes: QALE= Quality-Adjusted-Life-Expectancy

Over a lifetime, the intervention is associated with a positive net health quality impact and a consequent reduction of the baseline gap in QALE, by a 0.02%. In contrast, over the programme duration, the intervention is associated with a negative impact on health inequality in QALY terms, with a widening of the baseline gap by 0.002%. Considering a health equity impact plane (Cookson et al., 2017), the two time horizons present two opposite decision-making scenarios. A win-win scenario over a lifetime (North-East quadrant, intervention is cost-effective and improves health inequality) and a lose-lose scenario over 39-month time horizon (South-West quadrant, intervention is not cost-effective and harms health inequality).

6.3.4.2   Local Authority perspective

Figure 6.5 below compares the INMB values estimated from a health care and a Local Authority perspective, over a lifetime as well as a 39 month time horizon.

Figure 6.5 Impact of change in perspective on INMB
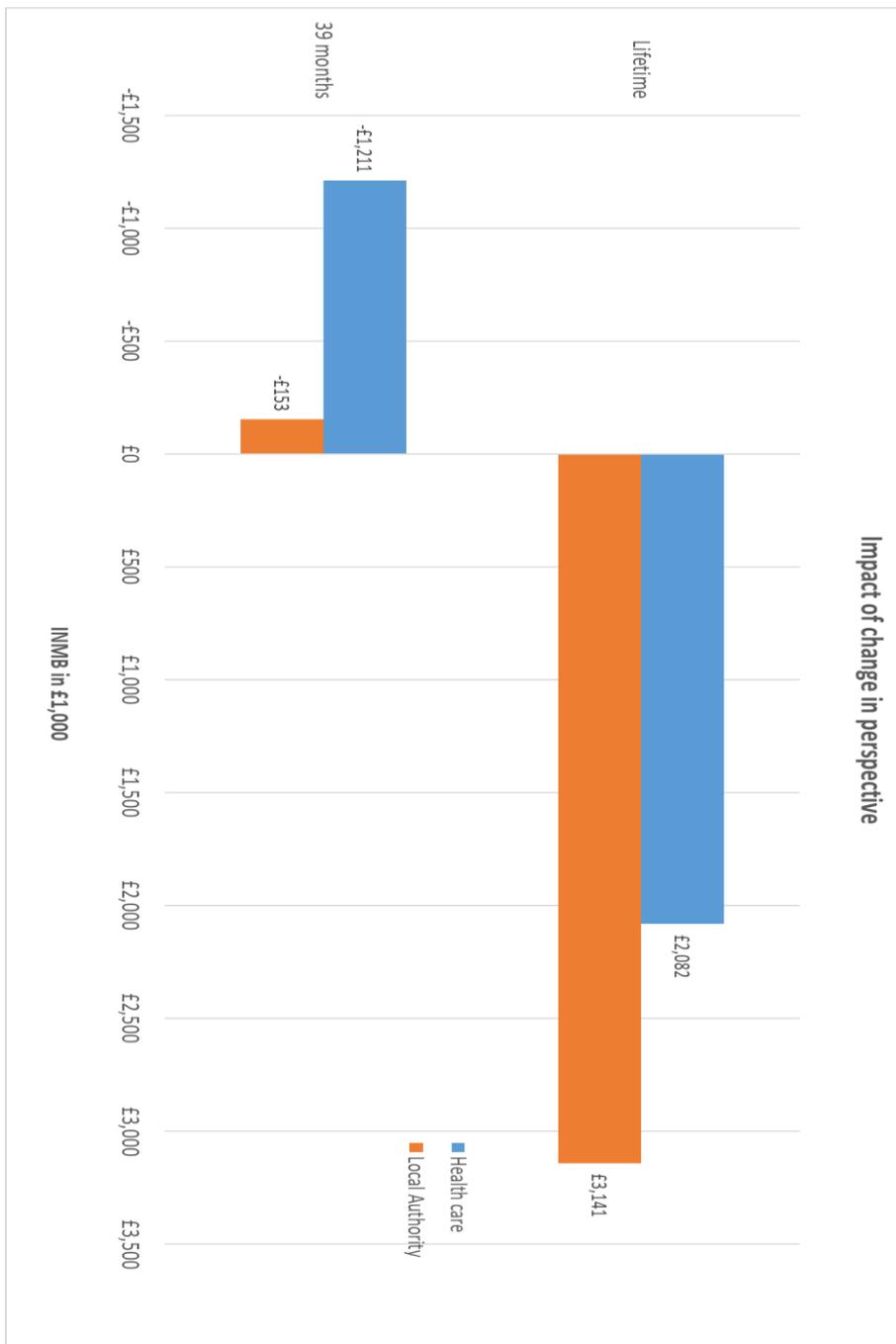
A positive difference of around £1 million was found when changing the perspective from a health care to that of a Local Authority, as the body administering the intervention. This difference is given by a differential per-participant programme cost of £20.41. No differential magnitude of change in INMB can be found between the two time horizons, as average approaches were used to calculate programme costs. However, this change in perspective results in a relatively small negative INMB (~£3 per participant), allowing LLGA almost to reach a break-even point in the short term.

6.4 Main findings

Overall, the simulated variations in modelling and methodological assumptions showed to impact costeffectiveness widely and variedly. The choice between methods of adjustments for selection bias and assumptions regarding the sustainability of intervention effects over time had vast implications in terms of projected INMBs, with potential to drive identification of the optimal alternative.

Choice of methods of adjustments for selection bias were shows to have significant implications in terms of projected INMBs. This was highlighted when choosing between a pragmatic (last observation carried forward, no intervention effect is not observed) and any of the formal approaches (complete case analysis, inverse probability weighing and Heckman selection model). Results also indicated how a relatively small average difference in PA behaviour can generate wide differences in health benefits at population-level over a lifetime.

Scenario results showed how, under current aggregate-level modelling approaches, different combinations of diseases can affect the number of cases averted, the projected incremental costs and health benefits (i.e. QALYs) in non-linear fashions. In particular, how the inclusion of additional diseases does not necessarily result in increases in the impact capture on an intervention, due to differential absolute risks, PA gradients and costs across selected chronic conditions.

Simulations showed that, under current modelling assumptions, short-term improvements in PA have the potential to produce long-term health benefits. This would occur even with improvements not sustained for more than six months, due to a large contribution made in terms of health utility gains by changes in PA levels. They also showed that assumptions regarding maintenance of behaviour change over time, and sub group heterogeneity, can influence population-level results widely, especially in the longer term.

In exploring alternative perspectives for EE, sub-group analyses showed how an average change in PA from individuals from deprived areas can contribute to population-level results, depending on differential risks of disease and their population proportions. Furthermore, findings indicated that, although deprived groups can potentially benefit more from equal improvement in PA on average, compared to non-deprived groups, negative inequality impacts can result especially in the short-term, if the health opportunity cost is taken into account. Finally, the choice of costing method used for estimation of

average programme costs can impact the economic results significantly, and make the difference between rendering the intervention cost-neutral or not.

6.5 Limitations

Missingness assumptions were addressed as mutually exclusive mechanisms of survey non-response (chapter 3). It was assumed that all of participants were surveyed a second time and, for each of three scenarios, participants' non-response was due to the same set of reasons. In fact, this was not likely to occur. For instance, differences between participants in terms of reasons for missingness might be due to the fact that follow-up surveys were, at least in part conducted, in person at the leisure centres, while around half of participants did not access them at all.

The identified assumptions were tested by means of scenario analysis, using the developed decision-analytic model, with estimates of relative cost-effectiveness being compared. Methods to address uncertainty, alternative to scenario analysis, are available in the health economic literature. In particular, some literature is available on incorporating structural uncertainty in value of information calculations. For example, Jackson et al. (2011) suggested a framework to formally incorporate structural uncertainty by inclusion of extra parameters within an expanded model, followed by model averaging. However, this approach inevitably requires judgement in regard to the choice of statistical methods to compute uncertainty and plausibility of alternative assumptions.

A model averaging approach was also proposed by Price et al. (2011) within an example of treatment for asthma, based on the idea of building and fitting a series of alternative models. Strong and Oakley (2014) illustrated a method that quantifies structural uncertainty by means a series of internal "discrepancy terms", of which the expected value of model improvement is calculated using VOI. Despite these methods being presented, they have received limited application and were not deemed fit for the purpose of this thesis. This is because the evaluative space of the model is unknown and application of those methods would have resulted in introducing further uncertainty in the analysis.

Testing of their impact on cost-effectiveness was not conducted in regard to the other relevant issues, such as that of distributional effectiveness and dose-response relationship between change in PA and health outcomes. This would have required building a series of parallel models and make assumptions regarding the comparability of input parameters between models, for each of these aspects. Balancing time and resources, it was decided not to focus on these assumptions.

Although the LLGA programme provided an example of a decision problem that is likely to repeat itself in other similar contexts, its data represented a major limitation in achieving an objective of illustrating the implications of variations in model assumptions.

The implications of assumptions regarding mechanisms of survey non-response may not have been accurately represented. In particular, the level of detail available on participants might not have sufficed to correct for selection bias, under a MAR, and particularly a MNAR scenario. Validity of these analyses was dependent on a correct model specification of the selection effects, and especially with the latter scenario, this might not be achieved due to a weak instrument. While being believed *a* priori to be correlated with the selection process, but not with the outcome (change in PA), the instrument was found to be statistically not independent of the outcome, possibly explaining the counterintuitive results.

However, the issue of identifying reliable instruments in MNAR settings is common (BaoLuo, 2016). Furthermore, these findings signified that, in highly constrained data settings where estimation is the main goal of analysis, pragmatic approaches may be preferred over principled and more sophisticated approaches to correct for selection bias.

Although the card swipe data provided a proxy measure of the decay of effect over time (i.e. programme attendance drop-off), allowing for testing the sensitivity of economic results to this key assumption, its validity could not be tested. While a proxy can risk accuracy of results, as often happens with population-level studies, such measures often remain the only option. In addition, in parallel to this research, the recent update of a nationally recognised off-the-shelf tool (Sport England, 2012) introduced a similar approach by assuming a rate of decay in drop-off participation over time.

To assess the health inequality implications of the intervention, neighbourhood-level deprivation was used as a proxy measure of socio-economic status. The IMD is a summary measure of area-level, rather than individual level deprivation. However, IMD score provides a nationally consistent measure that has been extensively used by local public health departments (Adams and White, 2006) and is a primary analytical tool for policy-makers.

A lack of data on leisure centre attendance outside LLGA, paid memberships, other sources of revenues and alternative uses of City Council resources limited the ability to reliably estimate the opportunity cost of LLGA programme from a Local Authority perspective. Estimation of this opportunity cost was based on two main assumptions. First, an implementation of the intervention under "steady-state" conditions, where no major structural variations, either to the hosting organisation or to current provision of activities could be envisaged. The second concerned the value lost from hosting such intervention, which was assumed to be equal to zero. While these represented relatively strong assumptions, they were made explicit and, considering the intervention nature, deemed likely to hold at least in the short term.

A lack of case study information and reliable data represented broader issues. Validity of results from any model is conditional on the quality of the input parameters used to populate it. For this reason, the results presented here need to be interpreted with caution. They remain surrounded by a

degree of uncertainty that is not possible to characterise (Mosleh, 1986), due to possible confounding effects on effectiveness and lack of appropriate knowledge of the data collection process, the implications of which could not be illustrated within this work.

Nevertheless, LLGA data served the purpose of testing key evaluation methods and assumptions that underpin previous economic models. Furthermore, results based on observational evidence is not an unusual situation in this setting. In fact, the level of information available for evaluation of the LLGA case study aligns with that of previous EEs (Cavill, 2011, Frew et al., 2014, Montes et al., 2012) and needs to be considered within an evaluation context that typifies non-research led, large scale programme of health promotion.

7 Chapter

Discussion and conclusions

This chapter summarises the preceding chapters and highlights the significance and limitations of the work generated in this thesis. This is structured according to how the thesis objectives were met. A summary of the lessons learnt to inform future evaluations of health promotion interventions and areas for further research follow, with final remarks concluding the chapter.

7.1 Overview of thesis findings

7.1.1 Review of the existing literature

To provide an overview of the current methodological challenges in EE of universal programmes to promote PA, a systematic review of existing EEs and models was conducted (chapter 2). This is the first comprehensive review proposing a critique of the analytical methods used for EE of PA promotion in the general population. A number of methodological gaps were identified and, in part, addressed by this thesis. One of the few in the growing field of health promotion, this review can be placed alongside previous methodological reviews aimed at highlighting areas for further research in public health evaluation (Alayli-Goebbels et al., 2014, Squires et al., 2016b, Weatherly et al., 2009).

7.1.1.1   Review update

To generate a current view of the available evidence, the literature search reported in chapter 2 was updated to include all studies published to mid-February 2019. It yielded 1306 articles, which were screened following the methods described in Chapter 2, of which five papers were selected. These included four EEs of universal strategies to promote PA in primary care (Gc et al., 2018), in school settings (Wang et al., 2017) and in the general population (Moore et al., 2017, Zapata-Diomedi et al., 2017). Gc et al. (2018), (Zapata-Diomedi et al., 2017) and (Verhoef et al., 2016) employed modelling techniques to assess the impact of PA interventions.

The methods used within these analyses were comparable to previous reviewed studies. Two studies focussed on children. Moore et al. (2017) estimated the effect of an incentive-based intervention employing a RCT design, while Wang et al. (2017) examined a multi-component programme aimed to increase PA during recess time in pupils, using a before-after approach.

Two other papers were model-based evaluations. Two modelling intervention scenarios of brief advice (Gc et al., 2018) and active transport (Zapata-Diomedi et al., 2017). The latter authors, comparably to Cobiac et al. (2009), used a multi-state life table approach in an Australian context. Gc et al. (2018) developed a discrete event simulation model, comparing three intervention modalities on a cohort of 10,000 representative adults of the English population. This analysis incorporated concerns regarding the maintenance of intervention effects, by simulation of alternative scenarios.

Verhoef et al. (2016) assessed the cost-effectiveness of a targeted programme offering free leisure centre membership to inactive individuals, over a lifetime horizon. They developed a simpler Markov model, compared to that by Frew et al. (2014) and (Roux et al., 2008), which was based on three levels of PA. In their base-case analysis, the effect of the intervention, which was assessed over a 4 month period, was assumed to last up to 12 months since the intervention started. With regard to perspective for EE, none of these analyses broadened the evaluation scope further, relative to the previously reviewed studies, or addressed equity implications formally.

7.1.2 Modelling methods development

Contributing to an overarching aim of developing analytical methods of EE, a novel approach to modelling the impact of universal programmes to promote healthy behaviours was devised (Chapter 4). A set of EMCs featuring a continuous-time mechanic is proposed as an integrated solution to the two modelling shortcomings identified in previous models, namely, population-level impact and maintenance of behaviour change over time.

The proposed modelling approach can be placed in a context of growing efforts to incorporate key concerns of public health policies into EE (Cookson et al., 2017, Squires et al. 2016a). It provides a flexible framework, which can be tailored to the context of any universal programme to promote healthy behaviours, maintaining a balance between complexity and practicality. While more sophisticated modelling solutions exist (e.g. individual-level modelling), this Excel-based aggregate-level approach may be intuitively more appealing to an audience of non-specialist modellers (e.g. public sector decision-

makers), to whom this model was targeted. However, a recent development of R software (R Shiny, @ Copyright 2017 RStudio Inc.) would also be a good fit for the purpose of a widespread use of the model in public health settings.

7.1.3 Implications of method choice for decision-making

The implications of method choice and respective assumptions regarding survey non-response, range of diseases, behaviour change maintenance over time and perspective for EE were illustrated in chapter 6. In response to recent reviews highlighting the importance of informative EEs, especially in a time of cuts to public health spending (Owen and Fischer, 2019, Owen et al., 2018), this analysis allows for a better understanding of how the identified modelling and methodological assumptions can affect optimal decision-making.

7.2 Contributions to the literature on universal interventions

In addressing the last thesis objective, a number of contributions to the literature on universal interventions of PA promotion have also been made throughout the thesis. While not being the focus of this thesis, the evaluation results generated here can be placed alongside those of previous similar studies.

The evaluation of LLGA programme has generated policy-relevant evidence relating to its intervention modality. In particular, evidence on the ability of

this programme to attract individuals from the resident population, especially

inactive and from low socio-economic areas, and its appeal to the target

audience. These findings, which were based on objective measurements (i.e.

card swipes), align with those of a recent work by Higgerson et al. (2018),

who used interrupted time series and difference-in-differences methods to

analyse leisure centre access and national survey data. Higgerson et al.

(2018) found implementation of the scheme ("Re:fresh") to be associated

with a greater proportion of those in the top 20% most deprived group

participating in leisure time PA (4.7%, 95% CI 4.4 to 5.0), compared to the

average population (3.9%, 95% CI 3.6 to 4.1). However, no EE of this

scheme was conducted or planned.

On the other hand, the results from the assessment of LLGA would have

been more credible if evaluation methods other than a simple before/after

approach were applied. Building on previous guidance (Craig et al. 2008), a

recent paper by Deidda et al. (2019) has suggested a framework for

conducting economic evaluations alongside natural experiments. However,

the ability to use more robust approaches was constrained by a retrospective

involvement into the project and lack of reliable external data that could have

been otherwise used to overcome the limitations of a before/after approach.

Only two of the reviewed EEs focussed on free access to leisure centre

activities (Frew et al., 2014, Vestergaard et al., 2006). As mentioned in

chapter 3.2, the study subject of EE by Frew et al. (2014) shared an

intervention modality similar to LLGA. The key difference with LLGA was that

Be Active offered the service in City Council-run leisure centres located not only in the most deprived city areas. For the remaining design choices, the two programmes were similar in their approach (e.g. universal offer of free access during off-peak times), decision making context (NHS and Local Authority) and population of interest (i.e. adult population of large city in England).

Verhoef et al. (2016) assessed the cost-effectiveness of a targeted strategy ("Give it a Go") dedicated to encourage PA participation in physically inactive individuals receiving state benefits by offering free four month leisure centre memberships in five facilities. However, unlike in LLGA, current gym members were excluded from participation, and participants had to attend a minimum of 5 times in order to qualify for the next month of free attendance, with a series of incentives being created to increase uptake.

To the best of my knowledge, the analysis conducted here represents the first empirical example of DCEA applied within the field of promotion of healthy behaviours. The distributional impact and cost-effectiveness of LLGA were explored, with scenario analyses based on changes in PA behaviour associated with implementation of the programme. While being exploratory, this evidence contributes to the broader and ongoing debate on universal versus targeted approaches to health promotion (Carey et al., 2015, Lorenc et al., 2013). This analysis can be placed within previous studies assessing the distributional impact of universal interventions (Asaria et al., 2016b, Dawkins et al., 2018). In addition, building on the analysis by Frew et al.

(2014), the economic cost of LLGA intervention was estimated from the perspective of Local Authorities, who may be interested in evaluating the impact of implementing this type of interventions in the future.

Finally, the issue of survey non-response was explored, within a nested survey approach. The case study provided an example of survey data collection process, which is imperfect, but likely to represent the sort of data collection that typifies this setting. A growing body of literature is concerned with the practical difficulties and the implications of assumptions regarding missing data mechanisms which, in health economics, has been mostly focused on RCT settings (Carpenter et al., 2002, Carpenter et al., 2007, Leurent et al., 2018a, Leurent et al., 2018b). In particular, this analysis can be placed alongside research efforts that have been spent on providing guidance for selection of suitable methods for dealing with missing data, and more recently, illustrating the implications of departures from common MCAR and MAR assumptions (Gomes et al., 2013, Gomes et al., 2019) .

7.3 Thesis scope

The methodological challenges addressed in this thesis are only part of the shortcomings characterising the EE of health promotion activities, and more broadly, public health interventions.The perspectives explored in the present analysis were limited in scope. Within a Local Authority perspective other possible intervention effects may have been relevant for inclusion. Alternative outcome measures of individual well-being, such as the capability measure

suggested by Nussbaum (1993), could have been used. However, the primary outcome, QALY, was designed to accumulate as a result of changes in PA states, and not only as mere consequences of reduced disease risks. Although the argument to look beyond a QALY is compelling as a common currency they provide a useful way of evaluating if these programmes are good value for money from a health care sector perspective. Without QALYs, it would be difficult to make a case in support or against.

Other aspects of social wellbeing may also have been relevant to include. For instance, social capital has emerged as an area of great interest by public policy makers (Rocco, 2012). While different definitions have been proposed in the literature, social capital is essentially concerned with the value of social participation and networks (Baum, 1999, Lynch et al., 2000). In the case study, participation to the programme, beyond possible changes in PA habits, might have generated relevant consequences, especially for marginalised and vulnerable groups (i.e. unemployed, ethnic minorities). Although challenges in the evidence of causality between social capital and individual's wellbeing has been acknowledged (Organisation for Economic Co-operation and Development, 2010), these may have represented relevant aspects to explore and capture for public health decision-making.

Taking a broader societal perspective offers opportunities to evaluate the impact of the intervention at a whole system level. Except for exploring different perspectives from alternative public sector agencies, the issue of inter-sectoral costs and consequences was not fully addressed. Previous

studies included environmental level impacts, such as reduction in traffic congestion and air pollution (Guo and Gandavarapu, 2010). However, relevance of these effects is dependent on the type of PA domain impacted by the intervention. In the case of a leisure centre based exercise offer, increased participation may have caused changes in transport patterns by participants (e.g. to go to the gym, instead of running in a park). Lack of data on participants' place of residence and means of transports ruled out the possibility of estimating these effects.

Advocated by Weatherly et al. (2009), a more comprehensive approach, such as general equilibrium modelling (The Scottish Government, 2016), would have allowed to capture effects beyond the public sector. For example, economic spill-overs in the private sector (e.g. leisure centre market) and out-of-pocket expenses (e.g. sport equipment, time, informal care), which were in part incorporated in previous models (Roux et al., 2015, Roux et al., 2008), could have been included.

Analysed results were not combined and provided as single outputs for decision-making. This was because of a lack of information about equity weights and decision-maker's preferences for the different objectives. To this purpose, a multi criteria decision analysis could be used to integrate the relevant evidence into one decision analysis tool and formal weighting of competing outcomes (Marsh et al., 2016, Thokala et al., 2016).

NICE recommend considering the use of CBA for EE of public health interventions. Stated preference approaches, such as contingent valuation methods, could have been used to estimate the value of non-health outcomes. In the paper by Frew et al. (2014), who assessed a similar intervention modality, this technique was used to estimate the value of the programme perceived by its participants. Due to contractual restrictions (i.e. Data Processing agreement) in respect of the possibility of contacting and collecting further data on participants, this could not be undertaken.

While the use of CBA presents some undeniable advantages, for instance, in that a single metric is used (i.e. monetary currency), application of this EE form has been limited in public health settings (Edwards et al., 2013, Weatherly et al., 2009). A more sophisticated approach, based on a CBA framework, is social return on investment (SROI). SROI mirrors that standard measures of financial return, but also allows for including societal values that are typically intangible and difficult to quantify (Social Value UK, 2012). Although this approach may be appealing for its ability to measure broader socio-economic outcomes and computing views of multiple stakeholders, the challenges related to the valuation of health states which characterise CBA methods have hampered its use also in public health settings (Banke-Thomas et al., 2015).

This method has been recommended by the UK Cabinet Office (The Cabinet Office, 2011) and allows for value beyond that of financial return to be captured. Nonetheless, decision-makers in public policy may not well receive

CBA methods (Bojke et al., 2018), due to lack of trust in monetised benefits (Hill et al., 2017).

7.4 Recommendations for design of future health promotion evaluations

7.4.1 Planning

The work conducted here afford the opportunity to provide a summary of the lessons learnt from the evaluation of the LLGA case study, which could be useful for future evaluations of similar programmes.

Prospective planning of the EE is key. An evaluation plan should be designed in the early stages, alongside programme design (Craig et al., 2008). As well as crucial aspects regarding the evaluation of the programme and involving the identification of behaviour, exposure / range of exposures, population of interest and outcome measures, the data collection process should reflect the question being asked.

When no RCT design is feasible, quasi-experimental options, such as natural experiment approaches should be implemented (Craig et al., 2012). Data on historic trends are important, especially if, as in the case of LLGA, the whole population is targeted. For instance, availability of data on leisure centre attendance during previous periods and in comparable populations would allow using time-series methods (e.g. interrupted time-series design) to

predict future trends in absence of the intervention and account for seasonal effects.

Timing and frequency of data collection will depend on how stable the reaction to the exposure is expected to be. In order to be able to estimate changes in trajectories, for instance through the use of latent grow modelling (Panter et al., 2017), at least two post-exposure assessments should be carried out. This would avoid relying on stronger assumptions of parallel trends.

If the endpoint of interest is improvement in health, rather than increase in leisure centre participation, the overall sphere of behaviour must be considered in the assessment of behaviour change. As a consequence, measures of overall PA, such as those used in LLGA might be used. However, attributing changes in leisure participation to observed changes in overall PA behaviour would require data on other PA domains (e.g. non-occupational), on their validity in the target population, and on possible confounders (e.g. changes in life circumstances) which may not be easy to detect and control for. Analysis of population-based longitudinal studies, such as that described by Lagerros et al. (2017), may be used to this purpose. However, country-related heterogeneity may limit their applicability to other contexts.

Availability of panel data, that is, of repeated observations on the same individuals followed up over multiple time points could improve the estimation

of effectiveness through more robust econometric approaches. If this method cannot be implemented, a repeated random sampling approach, where each individual in the population has an equal chance of being selected (i.e. random selection), would be the preferred option. If instead, for instance, interviewers go to the gyms to collect questionnaire data from present exercisers, it is possible that such measure of behaviour will not representative of the overall population (i.e. including that of non-exercisers). However, if gym attendance is measured and plausible to be the *only* reason for the difference between what measured and the "true" value, statistical methods (e.g. inverse probability weighting) can be used to correct for selection bias.

As addressed in the analysis of LLGA, selection effects can arise not only from the way individuals are identified for assessment. With survey measurements, the type of questionnaire used (and its validity) and the conditions in which individuals are asked to provide information can affect probability and level of response. For instance, characteristics of the interviewers (e.g. their experience), modality of administration (e.g. in person, self-administered, online), as well as personal characteristics of respondents (e.g. propensity to social desirability) can all have an influence on the measurement. For this reason, it is important to keep records of who collects the primary data and how they are collected, the number and characteristics of individuals asked to provide a measurement, and crucially, to collect information regarding the reasons for non-response by participants and for blank values (e.g. data not provided or lost in handling the data).

Data on individual characteristics are crucial for sub-group analysis and assessment of the distributional impact of interventions. In particular, as well as equity relevant characteristics (e.g. socio-economic status), access to data on place of residence (either or both home and work) would allow assessing whether, and to what extent, proximity plays a role in attracting and help changing behaviours by different individuals. While privacy concerns may arise in terms of possible identifications of individuals in the population (as with LLGA), anonymization techniques (e.g. data masking) allow us to reduce or eliminate the linkability of a dataset with the original identity of a participant (e.g. via an encryption scheme).

In order to build on previous efforts and experience, a portfolio of evaluation designs and reports should be collated. A collaboration network and sharing of expertise and data between Local Authorities would improve their ability to design and carry out evaluations of universal programmes, and by capitalising on existing knowledge, to avoid inefficiencies typical of start-up endeavours.

7.4.2 An economic perspective

The purpose of any EE is to solve an optimisation problem. In order to determine which of the competing and mutually exclusive alternatives provides the best value for money, a definition of the objective function is essential. This requires an a priori identification of the programme objectives, concerns and preferences by the decision-makers (e.g. for equity). Qualitative approaches, such as elicitation methods can be used for this purpose. While decision-makers may find it burdensome to provide

preference information (Janus et al., 2014), the lack of a clear definition leaves health economists with the choice of criteria upon which an intervention should be assessed, potentially leading to suboptimal decisions. To adequately inform decision-makers, the opportunity cost, rather than the financial cost of an option must be considered. Estimation of an opportunity cost is dependent on the identification of the competing alternatives which, if not correct, can lead to misleading results (i.e. extended dominance). Furthermore, given that the opportunity cost represents the value forgone from alternative actions, knowledge of the economic structure (e.g. organisational structure, budget life-cycles, current assets and their usage) of the entities affected by the decision would avoid relying on the related assumptions.

A solution would be to move from retrospective to prospective evaluations. Early involvement of a research team would be a useful support throughout all programme stages. While requiring more resources, this would mitigate some of the issues characterising these programmes. This would also allow for development of solutions to practical problems, such as privacy concerns and data handling, which limit the potential for valuable research outputs. Consolidation of collaborations with academic units would benefit both parties, contributing to the building of research and implementation capacity.

7.5 Areas for further research

There are many areas that future research could build on the work of this thesis. The proposed decision-analytic model should be further developed.

While the model was parameterised using the LLGA data and adapted to enable its EE, incorporation of five levels of socio-economic deprivation (i.e. IMD quintiles), would allow for a more granular assessment of the health inequality impact across the IMD distribution. Furthermore, aligning with previous analyses (Asaria et al., 2016b, Love-Koh et al., 2015), baseline population health distribution in terms of QALE at birth should be integrated into the model.

The model would benefit from relaxing the assumption of competing risk between disease states. A possible solution could be to adjust the probability of transition to the interacting disease state using model calibration methods (Taylor et al., 2010). However, to add this and other layers of complexity, the model could become unmanageable and transition from a spreadsheet (i.e. Excel) to programming language approach (e.g. R, MatLab or Python) may be required and has been suggested (Incerti et al., 2019).

The model could be used to assess the distributional impact and cost-effectiveness of other universal programmes. Thorough a collaboration with Sport England, Local Authorities and other research centres involved in previous evaluations, data on previous initiatives could be shared and a task force be formed with the aim of advancing our understanding of the health inequality implications of these programmes.

The present analysis did not include an evaluation of trade-offs between the two objectives of maximising population health and minimising unfair health

inequality. This was because under a lifetime or budget cycle time horizon, the intervention was not estimated to generate any trade-offs. Previous DCEAs used social welfare functions to represent inequality concerns, with a constant relative index of aversion to health inequity applied in all three of the published DCEAs (i.e. Atkinson), and a constant absolute gap index (i.e. Kolm) in one of these studies (Asaria et al., 2016b).

Aversion parameters can be estimated (Robson et al., 2017), or assumed to be used as reference points and calculate equally distributed equivalent levels for the health distribution. This allows to rank the strategies according to the level of social welfare produced. Further work should focus on assessing the distributional impact of re-design options of universal promotion of PA. This could help identify which strategy is optimal, and how alternative social value judgements influence this assessment.

More research on eliciting ranges of societal values for alternative distributions of health (i.e. absolute and relative levels inequality aversion and identification of what society deems as unfair variations in health) is required to support policymakers faced with real trade-offs between improving total health and reducing health inequality. For instance, a recent article by (Cookson et al., 2018) has proposed an e-learning approach for respondents faced with questions about trade-offs which may be hard to understand for lay people. This would contribute by defining a social welfare function, hence, enhancing the ability of health economists to generate an economic evidence that responds to society's demands.

When multiple private and public sector's budget constraints are affected by implementation of a health policy, an important question about how to apportion costs and benefits is still unanswered. This is not only empirically limited by a knowledge gap about the efficiency thresholds of public sectors, with that of health care likely to be lower than the commonly used £20,000 in the UK (Claxton et al., 2015), but also by a lack of analytical methods to incorporate the full range of impacts into an EE. To this respect, other than through a sector by sector approach, innovative methods for multi-sector analyses, such as one in the form of compensation test has been proposed (Claxton, 2007), but not tested empirically. More recently, Griffin et al. (2018) have proposed an analytical framework based on "impact inventories" to capture intervention's effects on different individuals, in terms of their opportunity costs.

7.6 Final remarks

The ultimate aim of any EE is to support decision-making, even when no high-quality data are available. EE of universal programmes to promote healthy behaviours can be challenging. Analysts face additional complexities which arise from lack of adequate research designs, imperfect knowledge of the behaviour change – population health and the decision-making processes. This leads to methodological simplifications and choices that can have wide implications for decision-making in public health.

This thesis has contributed to the understanding of these implications and proposed a simple and flexible modelling solution that overcomes some of

the current modelling shortcomings. More robust solutions will be possible when there are changes in research practice and a deeper knowledge of those processes is achieved. The value of an economic analysis, such that conducted within this PhD thesis, does not correspond to the degree of accuracy of its results. Its value lies between making an informed decision and enabling policy makers to deliberate on the allocation of public resources in an explicit and transparent manner.

References

ACADEMIC UNIT OF HEALTH ECONOMICS 2014. Checking for Duplicates Guidance. In: SPECIALISTS, A. U. O. H. E. I. (ed.).

ACHESON, E. D. 1988. On the State of the Public-Health - [the 4th Duncan Lecture]. Public Health, 102, 431-437.

ACTIVE LEEDS. 2013. Leeds Let's Get Active [Online]. Available: https://www.leedsletsgetactive.co.uk/ [Accessed 31.03.2019].

ADAMS, J. and WHITE, M., 2004. Why don't stage-based activity promotion interventions work? Health Education Research, 10.1093/her/cyg105.

ADAMS, J. & WHITE, M. 2006. Removing the health domain from the Index of Multiple Deprivation 2004-effect on measured inequalities in census measure of health. J Public Health (Oxf), 28, 379-83.

AJZEN, I. 1991. The Theory of Planned Behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.

ALASTAIR M. GRAY, P. M. C., JANE L. WOLSTENHOLME, SARAH WORDSWORTH 2011. Applied Methods of Cost-effectiveness Analysis in Healthcare, Oxford, Oxford University Press.

ALAYLI-GOEBBELS, A. F., EVERS, S. M., ALEXEEVA, D., AMENT, A. J., 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. J Public Health (Oxf), 36, 336-44.

ALLISON, R., BIRD, E. L. & MCCLEAN, S. 2017. Is Team Sport the Key to Getting Everybody Active, Every Day? A Systematic Review of Physical Activity Interventions Aimed at Increasing Girls' Participation in Team Sport. Aims Public Health, 4, 202-220.

AMARASINGHE, A. K. 2010. Cost-effectiveness implications of GP intervention to promote physical activity: evidence from Perth, Australia. Cost Eff Resour Alloc, 8, 10.

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.

ANSELIN, L. 1988. A Test for Spatial Auto-Correlation in Seemingly Unrelated Regressions. Economics Letters, 28, 335-341.

ASARIA, M., DORAN, T. & COOKSON, R. 2016a. The costs of inequality: whole-population modelling study of lifetime inpatient hospital costs in the English National Health Service by level of neighbourhood deprivation. J Epidemiol Community Health, 70, 990-6.

ASARIA, M., GRIFFIN, S. & COOKSON, R. 2016b. Distributional Cost-Effectiveness Analysis: A Tutorial. Medical Decision Making, 36, 8-19.

BABEY, S. H., WU, S. Y. & COHEN, D. 2014. How can schools help youth increase physical activity? An economic analysis comparing school-based programs. Preventive Medicine, 69, S55-S60.

BACHE, I. and REARDON, L. 2016. The Politics and Policy of Wellbeing. Understanding the Rise and Significance of a New Era. Cheltenham UK: Edward Elgar.

BAJEKAL, M., SCHOLES, S., LOVE, H., HAWKINS, N., O'FLAHERTY, M., RAINE, R. & CAPEWELL, S. 2012. Analysing Recent Socioeconomic Trends in Coronary Heart Disease Mortality in England, 2000-2007: A Population Modelling Study. Plos Medicine, 9.

BANDURA, A. 1985. Social foundations of thought and action: A social cognitive theory.

BANKE-THOMAS, A. O., MADAI, B., CHARLES, A., & VAN DEN BROEK, N. 2015. Social Return on Investment (SROI) methodology to account for value for money of public health interventions: a systematic review. BMC public health, 15, 582. doi:10.1186/s12889-015-1935-7

BAOLUO, S. 2016. Semi-Parametric Methods for Missing Data and Causal Inference. Harvard University.

BARRETT, J. L., GORTMAKER, S. L., LONG, M. W., WARD, Z. J., RESCH, S. C., MOODIE, M. L., CARTER, R., SACKS, G., SWINBURN, B. A., WANG, Y. C. & CRADOCK, A. L. 2015. Cost Effectiveness of an Elementary School Active Physical Education Policy. American Journal of Preventive Medicine, 49, 148-159.

BAUM, F. 1999. Social capital: is it good for your health? Issues for a public health agenda. Journal of Epidemiology and Community Health, 53, 195-196.

BEALE, S. J., BENDING, M. W., TRUEMAN, P. & NAIDOO, B. 2012. Should we invest in environmental interventions to encourage physical activity in England? An economic appraisal. European Journal of Public Health, 22, 869-873.

BENTHAM, J. 2000. An introduction to the Principles of Morals and Legislation, Kitchener.

BLOSSEY, R., CARDELLI, L. & PHILLIPS, A. 2006. A compositional approach to the stochastic dynamics of gene networks. Transactions on Computational Systems Biology Iv, 3939, 99-122.

BOJKE, L., SCHMITT, L., LOMAS, J., RICHARDSON, G. & WEATHERLY, H. 2018. Economic Evaluation of Environmental Interventions: Reflections on Methodological Challenges and Developments. International Journal of Environmental Research and Public Health, 15.

BOWSER, D., PATENAUDE, B., BHAWALKAR, M., DURAN, D. & BERMAN, P. 2019. Benefit incidence analysis in public health facilities in India: utilization and benefits at the national and state levels. International Journal for Equity in Health, 18.

BRAY, B. D., PALEY, L., HOFFMAN, A., JAMES, M., GOMPERTZ, P., WOLFE, C. D. A., HEMINGWAY, H., RUDD, A. G. & COLLABORATION, S. 2018. Socioeconomic disparities in first stroke incidence, quality of care, and survival: a nationwide registry-based cohort study of 44 million adults in England. Lancet Public Health, 3, E185-E193.

BRETTLE, A. J., LONG, A. F., GRANT, M. J. & GREENHALGH, J. 1998. Searching for information on outcomes: do you need to be comprehensive? Quality in Health Care, 7, 163-167.

BRIGGS, A., CLARK, T., WOLSTENHOLME, J. & CLARKE, P. 2003. Missing .... presumed at random: cost-analysis of incomplete data. Health Economics, 12, 377-392.

BRIGGS, A., SCULPHER, M., CLAXTON, K., 2006. Decision Modelling for Health Economic Evaluation, Oxford, Oxford University Press.

BRIGGS, A. D., WOLSTENHOLME, J., BLAKELY, T. & SCARBOROUGH, P. 2016. Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. Popul Health Metr, 14, 17.

BRIGGS, A. H. & GRAY, A. M. 1999. Handling uncertainty when performing economic evaluation of healthcare interventions. Health Technol Assess, 3, 1-134.

BROUWER, W. B. F., CULYER, A. J., VAN EXEL, N. J. A. & RUTTEN, F. F. H. 2008. Welfarism vs. extra-welfarism. Journal of Health Economics, 27, 325-338.

BRUG, J., CONNER, M., Harre, N., KREMERS, S., McKELLAR, S., WHITELAW, S., 2004. The Transtheoretical Model and stages of change: a critique: Observations by, [Online]. Available: https://doi.org/10.1093/her/cyh005

BUXTON, M. J., DRUMMOND, M. F., VAN HOUT, B. A., PRINCE, R. L., SHELDON, T. A., SZUCS, T. & VRAY, M. 1997. Modelling in economic evaluation: an unavoidable fact of life. Health Econ, 6, 217-27.

CANCER AUSTRALIA 2016. Bowel cancer statistics. [Online]. Available: https://bowel-cancer.canceraustralia.gov.au/statistics

CANCER RESEARCH UK 2011. Deprivation gradient for cancer incidence.

CAREY, G., CRAMMOND, B. & DE LEEUW, E. 2015. Towards health equity: a framework for the application of proportionate universalism. Int J Equity Health, 14, 81.

CARPENTER, J., POCOCK, S. & LAMM, C. J. 2002. Coping with missing data in clinical trials: a model-based approach applied to asthma trials. Stat Med, 21, 1043-66.

CARPENTER, J. R., KENWARD, M. G. & WHITE, I. R. 2007. Sensitivity analysis after multiple imputation under missing at random: a weighting approach. Stat Methods Med Res, 16, 259-75.

CAVILL, N., MULLER, L., MULHALL, C. & RUTTER, H. 2011. Cycling demonstration towns: A cost-effective investment to promote physical activity. Obesity Reviews, 12.

CENTRE FOR REVIEWS AND DISSEMINATION 2009. Systematic Reviews CRD's guidance for undertaking reviews in health care, CRD, University of York.

CHEN, I. J., CHOU, C. L., YU, S. & CHENG, S. P. 2008. Health services utilization and cost utility analysis of a walking program for residential community elderly. Nurs Econ, 26, 263-9.

CHIEF MEDICAL OFFICERS. 2019. UK Chief Medical Officers' Physical Activity Guidelines, United Kingdom. 2019 [Online]. Available:

References

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/832868/uk-chief-medical-officers-physical-activity-guidelines.pdf. [Accessed 6 November 2019].

CHRISTIANSEN, P., BJERRE, K., EJLERTSEN, B., JENSEN, M. B., RASMUSSEN, B. B., LAENKHOLM, A. V., KROMAN, N., EWERTZ, M., OFFERSEN, B., TOFTDAHL, D. B., MOLLER, S., MOURIDSEN, H. T. & GRP, D. B. C. C. 2011. Mortality Rates Among Early-Stage Hormone Receptor-Positive Breast Cancer Patients: A Population-Based Cohort Study in Denmark. Jnci-Journal of the National Cancer Institute, 103, 1363-1372.

CLARK, A. & ETILE, F. 2002. Do health changes affect smoking? Evidence from British panel data. Journal of Health Economics, 21, 533-562.

CLAVEL, J., MERCERON, G. & ESCARGUEL, G. 2014. Missing Data Estimation in Morphometrics: How Much is Too Much? Systematic Biology, 63, 203-218.

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.

CLAXTON, K., SCULPHER, M. AND CULYER T. 2007. Mark versus Luke? Appropriate Methods for the Evaluation of Public Health Interventions. In: YORK, U. O. (ed.) Woking Papers. York: University of York.

COAST, J., SMITH, R., LORGELLY, P.2008. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. Soc Sci Med. 2008, 67: 1190-1198. 10.1016/j.socscimed.2008.06.027.

COBIAC, L. J., VOS, T. & BARENDREGT, J. J. 2009. Cost-Effectiveness of Interventions to Promote Physical Activity: A Modelling Study. Plos Medicine, 6.

COCKERHAM, W. C. 2007. Social Causes of Health and Disease Cambridge, Polity Press

COOKSON, R., ALI, S., TSUCHIYA, A. & ASARIA, M. 2018. E-learning and health inequality aversion: A questionnaire experiment. Health Economics, 27, 1754-1771.

COOKSOM, R., McCABE, C., and TSUCHIYA, A. 2008. Public healthcare resource allocation and the Rule of Rescue. Journal of Medical Ethics, 34, 540-4

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

CRAIG, P., COOPER, C., GUNNELL, D., HAW, S., LAWSON, K., MACINTYRE, S., OGILVIE, D., PETTICREW, M., REEVES, B., SUTTON, M. & THOMPSON, S. 2012. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. Journal of Epidemiology and Community Health, 66, 1182-1186.

CRAIG, B., BUSSCHBACH, J. and SALOMON, J. 2009. Modeling Ranking, Time Trade-Off, and Visual Analog Scale Values for EQ-5D Health States. Medical Care, 47(6), pp.634-641.

CRAIG, C. L., MARSHALL, A. L., SJ??STR??M, M., BAUMAN, A. E., BOOTH, M. L., AINSWORTH, B. E., PRATT, M., EKELUND, U., YNGVE, A., SALLIS, J. F. AND OJA, P. 2000 International

Physical Activity Questionnaire: 12-Country Reliability and Validity. Medicine & Science in Sports & Exercise, 35(8), pp.1381-1395.

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, 337, a1655.

CUIJPERS, P., VOGELZANGS, N., TWISK, J., KLEIBOER, A., LI, J. & PENNINX, B. W. 2014. Comprehensive Meta-Analysis of Excess Mortality in Depression in the General Community Versus Patients With Specific Illnesses. American Journal of Psychiatry, 171, 453-462.

CULYER, A. J. 1990. Commodities, characteristics of commodities, charac-teristics of people, utilities, and the quality of life. In S. Baldwin (Ed.),Quality of life. Perspectives and policies(pp. 9–27). London: Routledge

CURTIS, E. M., VAN DER VELDE, R., MOON, R. J., VAN DEN BERGH, J. P. W., GEUSENS, P., DE VRIES, F., VAN STAA, T. P., COOPER, C. & HARVEY, N. C. 2016. Epidemiology of fractures in the United Kingdom 1988-2012: Variation with age, sex, geography, ethnicity and socioeconomic status. Bone, 87, 19-26.

DALLAT, M. A. T., SOERJOMATARAM, I., HUNTER, R. F., TULLY, M. A., CAIRNS, K. J. & KEE, F. 2014. Urban greenways have the potential to increase physical activity levels cost-effectively. European Journal of Public Health, 24, 190-195.

DAVIS, R., CAMPBELL, R., HILDON, Z., HOBBS, L. & MICHIE, S. 2015. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev, 9, 323-44.

DAWKINS, B. R., MIRELMAN, A. J., ASARIA, M., JOHANSSON, K. A. & COOKSON, R. A. 2018. Distributional cost-effectiveness analysis in low- and middle-income countries: illustrative example of rotavirus vaccination in Ethiopia. Health Policy Plan, 33, 456-463.

DE SMEDT, D., DE COCKER, K., ANNEMANS, L., DE BOURDEAUDHUIJ, I. & CARDON, G. 2012. A cost-effectiveness study of the community-based intervention '10 000 Steps Ghent'. Public Health Nutrition, 15, 442-451.

DEIDDA, M., GEUE, C., KREIF, N., DUNDAS, R. & MCINTOSH, E. 2019. A framework for conducting economic evaluations alongside natural experiments. Soc Sci Med, 220, 353-361.

DEPARTMENT FOR COMMUNITIES AND LOCAL GOVERNMENT 2016. The English Indices of Deprivation 2015 – Frequently Asked Questions (FAQs). In: GOVERNMENT, D. F. C. A. L. (ed.).

DEPARTMENT OF HEALTH 2011a. Physical activity guidelines for Adults (19–64 years). In: HEALTH, D. O. (ed.). [Online]. Available: https://www.nhs.uk/Livewell/fitness/Documents/adults-19-64-years.pdf

DEPARTMENT OF HEALTH 2011b. Physical activity guidelines for Adults (65+ years). In: HEALTH, D. O. (ed.). [Online]. Available: https://www.nhs.uk/Livewell/fitness/Documents/older-adults-65-years.pdf

DEPARTMENT OF HEALTH 2012. NHS Costing Manual. [Online]. Available: https://www.gov.uk/government/publications/nhs-costing-manual

DEVLIN, N. J., & BROOKS, R. 2017. EQ-5D and the EuroQol Group: Past, Present and Future. Applied health economics and health policy, 15(2), 127–137. https://doi.org/10.1007/s40258-017-0310-5

DING, D., LAWSON, K. D., KOLBE-ALEXANDER, T. L., FINKELSTEIN, E. A., KATZMARZYK, P. T., VAN MECHELEN, W., PRATT, M. & LANCET PHYSICAL ACTIVITY SERIES 2 EXECUTIVE, C. 2016. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. Lancet, 388, 1311-24.

DOLAN, P., 1997 Modelling valuations for EuroQol health states. Med Care 1997,35(11):1095–1108. 10.1097/00005650-199711000-00002

DOLAN, P., HALLSWORTH, MICHAEL, HALPERN, DAVID, KING, DOMINIC AND VLAEV, IVO 2010. MINDSPACE: influencing behaviour for public policy. In: GOVERNMENT, I. O. (ed.). London.

DORING, N., ZETHRAEUS, N., TYNELIUS, P., DE MUNTER, J., SONNTAG, D. & RASMUSSEN, F. 2018. Economic Evaluation of PRIMROSE - A Trial-Based Analysis of an Early Childhood Intervention to Prevent Obesity. Frontiers in Endocrinology, 9.

DOUC, R., MOULINES, E., PRIOURET, P. & SOULIER, P. 2018. Markov Chains.

DRUMMOND, M., SCULPHER, MJ, CLAXTON, K, STODDART, GL & TORRANCE 2015. Methods for the Economic Evaluation of Health Care Programmes, Oxford, Oxford University Press.

EDWARDS, R. T., CHARLES, J. M. & LLOYD-WILLIAMS, H. 2013. 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.

EPPI. 2019. CCEMG - EPPI-Centre Cost Converter [Online]. EPPI-Centre. Available: http://eppi.ioe.ac.uk/costconversion/default.aspx [Accessed].

ESPINO, J. N. V., GUERRERO, N., RHOADS, N., SIMON, N. J., ESCARON, A. L., MEINEN, A., NIETO, F. J. & MARTINEZ-DONATE, A. P. 2015. Community-Based Restaurant Interventions to Promote Healthy Eating: A Systematic Review. Preventing Chronic Disease, 12.

FARIA, R., GOMES, M., EPSTEIN, D. & WHITE, I. R. 2014. A Guide to Handling Missing Data in Cost-Effectiveness Analysis Conducted Within Randomised Controlled Trials. Pharmacoeconomics, 32, 1157-1170.

FENWICK, E., O'BRIEN, B. J. & BRIGGS, A. 2004. Cost-effectiveness acceptability curves--facts, fallacies and frequently asked questions. Health Econ, 13, 405-15.

FREW, E. J., BHATTI, M., WIN, K., SITCH, A., LYON, A., PALLAN, M. & ADAB, P. 2014. Cost-effectiveness of a community-based physical activity programme for adults (Be Active) in the UK: an economic analysis within a natural experiment. British Journal of Sports Medicine, 48, 207-212.

FRIED, L. P., TANGEN, C. M., WALSTON, J., NEWMAN, A. B., HIRSCH, C., GOTTDIENER, J., SEEMAN, T., TRACY, R., KOP, W. J., BURKE, G., MCBURNIE, M. A. & COLLABOR, C. H. S. 2001. Frailty in older adults: Evidence for a phenotype. Journals of Gerontology Series a-Biological Sciences and Medical Sciences, 56, M146-M156.

GARRETT, S., ELLEY, C. R., ROSE, S. B., O'DEA, D., LAWTON, B. A. & DOWELL, A. C. 2011. Are physical activity interventions in primary care and the community cost-effective? A systematic review of the evidence. Br J Gen Pract, 61, e125-33.

GC, V., WILSON, E. C., SUHRCKE, M., HARDEMAN, W., SUTTON, S. & TEAM, V. B. I. P. 2016. Are brief interventions to increase physical activity cost-effective? A systematic review. Br J Sports Med, 50, 408-17.

GC, V. S., SUHRCKE, M., HARDEMAN, W., SUTTON, S., WILSON, E. C. F. & VERY BRIEF INTERVENTIONS PROGRAMME, T. 2018. Cost-Effectiveness and Value of Information Analysis of Brief Interventions to Promote Physical Activity in Primary Care. Value Health, 21, 18-26.

GIOVANNUCCI, E., HARLAN, D. M., ARCHER, M. C., BERGENSTAL, R. M., GAPSTUR, S. M., HABEL, L. A., POLLAK, M., REGENSTEINER, J. G. & YEE, D. 2010. Diabetes and Cancer: A Consensus Report. Ca-a Cancer Journal for Clinicians, 60, 207-221.

GOLSTEIJN, R. H. J., PEELS, D. A., EVERS, S. M. A. A., BOLMAN, C., MUDDE, A. N., DE VRIES, H. & LECHNER, L. 2014. Cost-effectiveness and cost-utility of a Web-based or print-delivered tailored intervention to promote physical activity among adults aged over fifty: an economic evaluation of the Active Plus intervention. International Journal of Behavioral Nutrition and Physical Activity, 11.

GOMES, M., DIAZ-ORDAZ, K., GRIEVE, R. & KENWARD, M. G. 2013. Multiple Imputation Methods for Handling Missing Data in Cost-effectiveness Analyses That Use Data from Hierarchical Studies: An Application to Cluster Randomized Trials. Medical Decision Making, 33, 1051-1063.

GOMES, M., RADICE, R., CAMARENA BRENES, J. & MARRA, G. 2019. Copula selection models for non-Gaussian outcomes that are missing not at random. Stat Med, 38, 480-496.

GORDON, L., GRAVES, N., HAWKES, A. & EAKIN, E. 2007. A review of the cost-effectiveness of face-to-face behavioural interventions for smoking, physical activity, diet and alcohol. Chronic Illn, 3, 101-29.

GOYDER, E., HIND, D., BRECKON, J., DIMAIRO, M., MINTON, J., EVERSON-HOCK, E., READ, S., COPELAND, R., CRANK, H., HORSPOOL, K., HUMPHREYS, L., HUTCHISON, A., KESTERTON, S., LATIMER, N., SCOTT, E., SWAILE, P., WALTERS, S. J., WOOD, R., COLLINS, K. & COOPER, C. 2014. A randomised controlled trial and cost-effectiveness evaluation of 'booster' interventions to sustain increases in physical activity in middle-aged adults in deprived urban neighbourhoods. Health Technology Assessment, 18, 1-+.

GRIFFIN, S., WALKER, S., SCULPHER, M. & ASARIA, M. 2018. An Analytical Framework for Economic Evaluation of Interventions with Effects on Multiple Outcomes, Costs Falling on Different Budgets, and Involving More Than One Decision Maker. Value in Health, 21, S10-S10.

GRIFFITHS, U. K., ANIGBOGU, B. & NANCHAHAL, K. 2012. Economic evaluations of adult weight management interventions: a systematic literature review focusing on methods used for determining health impacts. Appl Health Econ Health Policy, 10, 145-62.

GROESSL, E. J., KAPLAN, R. M., SWEET, C. M. C., CHURCH, T., ESPELAND, M. A., GILL, T. M., GLYNN, N. W., KING, A. C., KRITCHEVSKY, S., MANINI, T., MCDERMOTT, M. M., REID, K. F.,

References

RUSHING, J., PAHOR, M. & GRP, L. S. 2016. Cost-effectiveness of the LIFE Physical Activity Intervention for Older Adults at Increased Risk for Mobility Disability. Journals of Gerontology Series a-Biological Sciences and Medical Sciences, 71, 656-662.

GROSSMAN, M. 1972. The Demand for Health: A Theoretical and Empirical Investigation.

GULLIFORD, M., CHARLTON, J., BHATTARAI, N. & RUDISILL, C. 2014. Social and material deprivation and the cost-effectiveness of an intervention to promote physical activity: cohort study and Markov model. Journal of Public Health, 36, 674-683.

GUO, J. Y. & GANDAVARAPU, S. 2010. An economic evaluation of health-promotive built environment changes. Prev Med, 50 Suppl 1, S44-9.

GUSI, N., REYES, M. C., GONZALEZ-GUERRERO, J. L., HERRERA, E. & GARCIA, J. M. 2008. Cost-utility of a walking programme for moderately depressed, obese, or overweight elderly women in primary care: a randomised controlled trial. BMC Public Health, 8, 231.

HAAS, M. 2006. Economic analysis of tai chi as a means of preventing falls and falls related injuries among older adults. NHS Economic Evaluation Database (NHSEED).

HART, J. T. 1971. The inverse care law. Lancet, 1, 405-12.

HENDERSON, J. W. 2014. Health Economics and Policy.

HERTZMAN, C. S., ARJUMAND A. 2009. Successful Societies: How Institutions and Culture Affect Health. In: PETER A. & LAMONT, M. (ed.) Population Health and the Dynamics of Collective Development. Cambridge: Mass.: Cambridge University Press.

HIGGERSON, J., HALLIDAY, E., ORTIZ-NUNEZ, A., BROWN, R. & BARR, B. 2018. Impact of free access to leisure facilities and community outreach on inequalities in physical activity: a quasi-experimental study. Journal of Epidemiology and Community Health, 72, 252-258.

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 2018. The Green Book. Central Government Guidance on Appraisal and Evaluation. London: OGL. [Online] Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/685903/The_Green_Book.pdf

HOCH, J. S., BRIGGS, A. H. & WILLAN, A. R. 2002. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ, 11, 415-30.

HOLDEN, J., and HARDING, A. 2015. Using evidence: Great Manchester case study. (2015). What Works Centre for Local Economic Growth, Liverpool [Online] Available at: https://whatworksgrowth.org/public/files/Best_Practise_studies/15-08-04_Using_evidence_-_Greater_Manchester_Case_Study.pdf [Accessed 12 Feb. 2020].

HUSEREAU D., DRUMMOND, M., PETROU, S., CARSWELL. C., MOHER, D., GREENBERG, D., 2013 Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement BMJ ; 346 :f1049 https://www.bmj.com/content/346/bmj.f1049

INCERTI, D., THOM, H., BAIO, G. & JANSEN, J. P. 2019. R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. Value Health, 22, 575-579.

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.

JACKSON, C. H., BOJKE, L., THOMPSON, S. G., CLAXTON, K. & SHARPLES, L. D. 2011. A framework for addressing structural uncertainty in decision models. Med Decis Making, 31, 662-74.

JACOBS-VAN DER BRUGGEN, M. A. M., BOS, G., BEMELMANS, W. J., HOOGENVEEN, R. T., VIJGEN, S. M. & BAAN, C. A. 2007. Lifestyle interventions are cost-effective in people with different levels of diabetes risk - Results from a modeling study. Diabetes Care, 30, 128-134.

JAGO, R., BARANOWSKI, T., BARANOWSKI, J. C., CULLEN, K. W. & THOMPSON, D. I. 2007. Social desirability is associated with some physical activity, psychosocial variables and sedentary behavior but not self-reported physical activity among adolescent males. Health Education Research, 22, 438-449.

JAMES LOVE-KOH, J., PENNINGTON, R., OWEN, L. & GRIFFIN, S. 2017. Frameworks for considering health inequalities and their suitability for use within NICE public health guidelines.

JANUS, S. I. M., WEERNINK, M. G. M., VAN TIL, J. A., RAISCH, D. W., VAN MANEN, J. G. & IJZERMAN, M. J. 2014. A Systematic Review to Identify the Use of Preference Elicitation Methods in Health Care Decision Making. Value in Health, 17, A515-A515.

JOANNA BRIGGS INSTITUTE 2017. Joanna Briggs Institute Reviewer's Manual. In: INSTITUTE, T. J. B. (ed.). Adelaide: University of Adelaide.

JONES, A. M. 2007. Applied Econometrics for Health Economists: A Practical Guide.

JOSEPH, J., SVARTBERG, J., NJOLSTAD, I. & SCHIRMER, H. 2010. Incidence of and risk factors for type-2 diabetes in a general population: The Tromso Study. Scandinavian Journal of Public Health, 38, 768-775.

KRIEGER, N. 2001. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol, 30, 668-77.

KULMALA, J., NYKANEN, I. & HARTIKAINEN, S. 2014. Frailty as a predictor of all-cause mortality in older men and women. Geriatrics & Gerontology International, 14, 899-905.

LAGERROS, Y. T., HANTIKAINEN, E., MARIOSA, D., YE, W. M., ADAMI, H. O., GROTTA, A., GHILOTTI, F. & BELLOCCO, R. 2017. Cohort Profile: The Swedish National March Cohort. International Journal of Epidemiology, 46, 795-+.

LAINE, J., KUVAJA-KOLLNER, V., PIETILA, E., KOIVUNEVA, M., VALTONEN, H. & KANKAANPAA, E. 2014. Cost-effectiveness of population-level physical activity interventions: a systematic review. Am J Health Promot, 29, 71-80.

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. Int J Behav Nutr Phys Act, 12, 140.

LAST, J., A dictionary of epidemiology. New York: Oxford University Press,2001.

LEE, P. H., MACFARLANE, D. J., LAM, T. H., & STEWART, S. M. 2011. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. The international journal of behavioral nutrition and physical activity, 8, 115. doi:10.1186/1479-5868-8-115

LEE, I. M., SHIROMA, E. J., LOBELO, F., PUSKA, P., BLAIR, S. N., KATZMARZYK, P. T., & Lancet Physical Activity Series Working Group 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet (London, England), 380(9838), 219–229. doi:10.1016/S0140-6736(12)61031-9

LEE, I. M., SHIROMA, E. J., LOBELO, F., PUSKA, P., BLAIR, S. N., KATZMARZYK, P. T. & WORKIN, L. P. A. S. 2012. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet, 380, 219-229.

LEURENT, B., GOMES, M. & CARPENTER, J. R. 2018a. Missing data in trial-based cost-effectiveness analysis: An incomplete journey. Health Econ, 27, 1024-1040.

LEURENT, B., GOMES, M., FARIA, R., MORRIS, S., GRIEVE, R. & CARPENTER, J. 2018b. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis Using Multiple Imputation. Value in Health, 21, S14-S14.

LIN, C. C., LI, C. I., CHANG, C. K., LIU, C. S., LIN, C. H., MENG, N. H., LEE, Y. D., CHEN, F. N. & LI, T. C. 2011. Reduced Health-Related Quality of Life in Elders with Frailty: A Cross-Sectional Study of Community-Dwelling Elders in Taiwan. Plos One, 6.

LOCAL PARTNERSHIPS 2014. Commissioning in Local Government: a research project for Local Partnerships.

LORENC, T., PETTICREW, M., WELCH, V. & TUGWELL, P. 2013. What types of interventions generate inequalities? Evidence from systematic reviews. Journal of Epidemiology and Community Health, 67, 190-193.

LOVE-KOH, J., ASARIA, M., COOKSON, R. & GRIFFIN, S. 2015. The Social Distribution of Health: Estimating Quality-Adjusted Life Expectancy in England. Value Health, 18, 655-62.

LUENGO-FERNANDEZ, R., GRAY, A. M., ROTHWELL, P. M. & OXFORD VASCULAR, S. 2012. A population-based study of hospital care costs during 5 years after transient ischemic attack and stroke. Stroke, 43, 3343-51.

LYNCH, J., DUE, P., MUNTANER, C. & SMITH, G. D. 2000. Social capital - Is it a good investment strategy for public health? Journal of Epidemiology and Community Health, 54, 404-408.

MA, J. S., CHAN, W. & TILLEY, B. C. 2018. Continuous time Markov chain approaches for analyzing transtheoretical models of health behavioral change: A case study and comparison of model estimations. Statistical Methods in Medical Research, 27, 593-607.

MAHESWARAN, H., PETROU, S., REES, K. & STRANGES, S. 2013. Estimating EQ-5D utility values for major health behavioural risk factors in England. J Epidemiol Community Health, 67, 172-80.

MARMOT, M. 2001. From Black to Acheson: two decades of concern with inequalities in health. A celebration of the 90th birthday of Professor Jerry Morris. International Journal of Epidemiology, 30, 1165-1171.

MARMOT, M. 2005. Social determinants of health inequalities. Lancet, 365, 1099-1104.

MARMOT, M. 2010. Fair Society, Healthy Lives The Marmot Review.

MARMOT, M. G., SMITH, G. D., STANSFELD, S., PATEL, C., NORTH, F., HEAD, J., WHITE, I., BRUNNER, E. & FEENEY, A. 1991. Health Inequalities among British Civil-Servants - the Whitehall-Ii Study. Lancet, 337, 1387-1393.

MARSH, K., IJZERMAN, M., THOKALA, P., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONNGREN, T., MUSSEN, F., PEACOCK, S., WATKINS, J. & DEVLIN, N. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making-Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value in Health, 19, 125-137.

MASTERS, R., ANWAR, E., COLLINS, B., et al 2017. Return on investment of public health interventions: a systematic review J Epidemiol Community Health 2017;71:827-834.

 MCCARTHY, R. 2006. On the dynamics of health capital accumulation. Soc Sci Med, 63, 817-28.

MCEACHAN, R. R., LAWTON, R. J., JACKSON, C., CONNER, M., MEADS, D. M. & WEST, R. M. 2011. Testing a workplace physical activity intervention: a cluster randomized controlled trial. Int J Behav Nutr Phys Act, 8, 29.

McGOVERN DPB, VALORI RM, SUMMERSKILL WSM, eds. Key topics in evidence based medicine. Oxford: BIOS Scientific Publishers,2001:26–9.

McINTOSH, E., CLARKE, P., FREW, E., LOUVIERE, J., 2010. Applied Methods of Cost-Benefit Analysis in Health Care, Oxford, Oxford University Press.

MCNAMEE, P., GREGSON, B. A., BUCK, D., BAMFORD, C. H., BOND, J. & WRIGHT, K. 1999. Costs of formal care for frail older people in England: the resource implications study of the MRC cognitive function and ageing study (RIS MRC CFAS). Social Science & Medicine, 48, 331-341.

MCPHEE, J. S., FRENCH, D. P., JACKSON, D., NAZROO, J., PENDLETON, N. & DEGENS, H. 2016. Physical activity in older age: perspectives for healthy ageing and frailty. Biogerontology, 17, 567-580.

MECHANIC, D. 2002. Disadvantage, inequality, and social policy. Health Affairs, 21, 48-59.

MEDICAL RESEARCH COUNCIL 2010. Report of a joint MRC Population Health Sciences Research Network/MRC Methodology Research Panel Workshop: using natural experiments to evaluate population health interventions. London: King's Fund.

MENG, X. F. & D'ARCY, C. 2013. The projected effect of increasing physical activity on reducing the prevalence of common mental disorders among Canadian men and women: A national population-based community study. Preventive Medicine, 56, 59-63.

MHOON, K. B., CHAN, W., DEL JUNCO, D. J. & VERNON, S. W. 2010. A Continuous-Time Markov Chain Approach Analyzing the Stages of Change Construct from a Health Promotion Intervention. JP J Biostat, 4, 213-226.

MICHIE, S., VAN STRALEN, M. M. & WEST, R. 2011. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci, 6, 42.

MIKKELSEN, M. V., HUSBY, S., SKOV, L. R. & PEREZ-CUETO, F. J. 2014. A systematic review of types of healthy eating interventions in preschools. Nutr J, 13, 56.

MOFFIT, R., FITZGERALD, J. & GOTTSCHALK, P. 1999. Sample Attrition in Panel Data: The Role of Selection on Observable. Annales d'Économie et de Statistique, 55/56, pp. 129-152.

MONTES, F., SARMIENTO, O. L., ZARAMA, R., PRATT, M., WANG, G. J., JACOBY, E., SCHMID, T. L., RAMOS, M., RUIZ, O., VARGAS, O., MICHEL, G., ZIEFF, S. G., VALDIVIA, J. A., CAVILL, N. & KAHLMEIER, S. 2012. Do Health Benefits Outweigh the Costs of Mass Recreational Programs? An Economic Analysis of Four Ciclovia Programs. Journal of Urban Health-Bulletin of the New York Academy of Medicine, 89, 153-170.

MOODIE, M., HABY, M., GALVIN, L., SWINBURN, B. & CARTER, R. 2009. Cost-effectiveness of active transport for primary school children - Walking School Bus program. International Journal of Behavioral Nutrition and Physical Activity, 6.

MOORE, J. B., HEBOYAN, V., ONIFFREY, T. M., BRINKLEY, J., ANDREWS, S. M. & KOLBE, M. B. 2017. Cost-effectiveness of Community-Based Minigrants to Increase Physical Activity in Youth. Journal of Public Health Management & Practice, 23, 364-369.

MOSLEH, A. 1986. Hidden Sources of Uncertainty - Judgment in the Collection and Analysis of Data. Nuclear Engineering and Design, 93, 187-198.

MULLER-RIEMENSCHNEIDER, F., REINHOLD, T. & WILLICH, S. N. 2009. Cost-effectiveness of interventions promoting physical activity. Br J Sports Med, 43, 70-6.

MUNRO, J. F., NICHOLL, J. P., BRAZIER, J. E., DAVEY, R. & COCHRANE, T. 2004. Cost effectiveness of a community based exercise programme in over 65 year olds: cluster randomised trial. Journal of Epidemiology and Community Health, 58, 1004-1010.

NATIONAL CAREERS SERVICE 2018. Personal trainer. In: GOVERNMENT, U. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2007. Behaviour change: general approaches. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2010. Smoking prevention in schools. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2012. Methods for the development of NICE public health guidance (third edition). In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2013. Guide to the methods of technology appraisal 2013. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2014. Behaviour change: individual approaches. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2015. Resource impact statement: Depression and anxiety disorder.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2016. Community engagement: improving health and wellbeing and reducing health inequalities. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE 2018. Developing NICE guidelines: the manual. In: EXCELLENCE, N. I. F. H. A. C. (ed.).

NHS SCOTLAND, 2016. Economics of Prevention [Online]. Available: http://www.healthscotland.scot/media/1089/economics-of-prevention-mar16.pdf [Accessed 10 November 2019]

NIEDERDEPPE, J., LLIANG, X., CROCK, B. & SKELTON, A. 2008. Media campaigns to promote smoking cessation among socioeconomically disadvantaged populations: What do we know, what do we need to learn, and what should we do now? Social Science & Medicine, 67, 1343-1355.

NSHIMYUMUKIZA, L., DURAND, A., GAGNON, M., DOUVILLE, X., MORIN, S., LINDSAY, C., DUPLANTIE, J., GAGNE, C., JEAN, S., GIGUERE, Y., DODIN, S., ROUSSEAU, F. & REINHARZ, D. 2013. An economic evaluation: Simulation of the cost-effectiveness and cost-utility of universal prevention strategies against osteoporosis-related fractures. J Bone Miner Res, 28, 383-94.

NUSSBAUM, M. A. S. A. 1993. The Quality of Life, Oxford, Oxford University Press.

NUTBEAM, D. 1998. Evaluating health promotion - progress, problems and solutions. Health Promotion International, 13, 27-44.

NWANERI, C., BOWEN-JONES, D., COOPER, H., CHIKKAVEERAPPA, K. & AFOLABI, B. A. 2012. Falling mortality rates in Type 2 diabetes mellitus in the Wirral Peninsula: a longitudinal and retrospective cohort population-based study. Postgraduate Medical Journal, 88, 679-683.

OFFICE OF DISEASE PREVENTION AND HEALTH PROMOTION 2018. Physical Activity Guidelines Advisory Committee Scientific Report In: SERVICES, U. S. D. O. H. A. H. (ed.).

OLSSON, T. (2011). Comparing top-down and bottom-up costing approaches for economic evaluation within social welfare. The European Journal of Health Economics, 12(5), 445-453. Retrieved from www.jstor.org/stable/41474381

O'NEIL, M. E., FRAGALA-PINKHAM, M. A., FORMAN, J. L. & TROST, S. G. 2014. Measuring reliability and validity of the ActiGraph GT3X accelerometer for children with cerebral palsy: a feasibility study. J Pediatr Rehabil Med, 7, 233-40.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT 2010. Social capital, human capital and health: What is the evidence? : Centre for Educational Research and Innovation.

OVER, E. A., WENDEL-VOS, G. W., VAN DEN BERG, M., REENEN, H. H., TARIQ, L., HOOGENVEEN, R. T. & VAN BAAL, P. H. 2012. Cost-effectiveness of counseling and pedometer use to increase physical activity in the Netherlands: a modeling study. Cost Eff Resour Alloc, 10, 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. 2018. The cost-effectiveness of public health interventions examined by NICE from 2011 to 2016. Journal of Public Health, 40, 557-566.

PANTER, J., OGILVIE, D. & CONSORTIUM, I. 2017. Can environmental improvement change the population distribution of walking? Journal of Epidemiology and Community Health, 71, 528-535.

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. 2011. The clinical effectiveness and cost-effectiveness of exercise referral schemes: a systematic review and economic evaluation. Health Technol Assess, 15, i-xii, 1-254.

PETERSON, M., CHANDLEE, M. & ABRAHAM, A. 2008. Cost-effectiveness analysis of a statewide media campaign to promote adolescent physical activity. Health Promot Pract, 9, 426-33.

PETRELLA, R. 2006. Cost effectiveness of a community-based exercise program for older adults. Clin J Sport Med, 16, 191-3.

PHILIPS, C., 2009. What is a QALY? What is...? series. Hayward Medical Communications. Available from http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/qaly.pdf [Accessed 08/03/2016]

PHILIPS, Z., BOJKE, L., SCULPHER, M., CLAXTON, K. & GOLDER, S. 2006. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. Pharmacoeconomics, 24, 355-71.

PHYSICAL ACTIVITY GUIDELINED ADVISORY COMMITTEE. 2018. Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services, 2018

POKHREL. S., 2015. Return on investment (ROI) modelling in public health: strengths and limitations, European Journal of Public Health, Volume 25, Issue 6, December 2015, Pages 908–909, https://doi.org/10.1093/eurpub/ckv136

POPULATION HEALTH SCIENCE RESEARCH NETWORK 2012. Methods for the economic evaluation of population health interventions: conceptual and practical challenges Glasgow.

PRICE, M. J., WELTON, N. J., BRIGGS, A. H. & ADES, A. E. 2011. Model averaging in the presence of structural uncertainty about treatment effects: influence on treatment decision and expected value of information. Value Health, 14, 205-18.

PRINGLE, A., COOKE, C., GILSON, N., MARSH, K. & MCKENNA, J. 2010. Cost-effectiveness of interventions to improve moderate physical activity: A study in nine UK sites. Health Education Journal, 69, 211-224.

PROCHASKA, J. O. & VELICER, W. F. 1997. The transtheoretical model of health behavior change. American Journal of Health Promotion, 12, 38-48. REVIEWS, C. A. S. 2019. Levels of evidence.

ROBEYNS, I. 2016. The Capability Approach, The Stanford Encyclopedia of Philosophy, Winter Edition, Edward N. Zalta (ed.), [Online] Available at https://plato.stanford.edu/archives/win2016/entries/capability-approach/

ROBSON, M., ASARIA, M., COOKSON, R., TSUCHIYA, A. & ALI, S. 2017. Eliciting the Level of Health Inequality Aversion in England. Health Economics, 26, 1328-1334.

ROCCO, L. A. S., M. 2012. Is social capital good for health? A European perspective. Copenhagen, : WHO Regional Office for Europe.

RODRIGUEZ-MARTINEZ, C. E., SOSSA-BRICENO, M. P. & CASTRO-RODRIGUEZ, J. A. 2018. A cost-effectiveness threshold analysis of a multidisciplinary structured educational intervention in pediatric asthma. J Asthma, 55, 561-570.

ROUX, L., PRATT, M., LEE, I. M., BAZZARRE, T. & BUCHNER, D. 2015. Does age modify the cost-effectiveness of community-based physical activity interventions? J Phys Act Health, 12, 224-31.

ROUX, L., PRATT, M., TENGS, T. O., YORE, M. M., YANAGAWA, T. L., VAN DEN BOS, J., RUTT, C., BROWNSON, R. C., POWELL, K. E., HEATH, G., KOHL, H. W., 3RD, TEUTSCH, S., CAWLEY, J., LEE, I. M., WEST, L. & BUCHNER, D. M. 2008. Cost effectiveness of community-based physical activity interventions. Am J Prev Med, 35, 578-88.

SEAMAN, S. R. & WHITE, I. R. 2013. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res, 22, 278-95.

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 adults - Results of project ACTIVE. American Journal of Preventive Medicine, 19, 1-8.

SHARMA, M., NAZARETH, I. & PETERSENT, I. 2016. Trends in incidence, prevalence and prescribing in type 2 diabetes mellitus between 2000 and 2013 in primary care: a retrospective cohort study. Bmj Open, 6.

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.

SILFEE, V. J., HAUGHTON, C. F., JAKE-SCHOFFMAN, D. E., LOPEZ-CEPERO, A., MAY, C. N., SREEDHARA, M., ROSAL, M. C. & LEMON, S. C. 2018. Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. Prev Med Rep, 11, 74-80.

SKOCPOL, T. 1991. Targeting within Universalism - Politically Viable Policies to Combat Poverty in the United-States. Urban Underclass, 411-436.

SPORT ENGLAND. 2012. Moves tool [Online]. Sport England. Available: https://www.sportengland.org/our-work/health-and-inactivity/what-is-moves/ [Accessed].

SQUIRES, H., CHILCOTT, J., AKEHURST, R., BURR, J. & KELLY, M. P. 2016a. A Framework for Developing the Structure of Public Health Economic Models. Value in Health, 19, 588-601.

SQUIRES, H., CHILCOTT, J., AKEHURST, R., BURR, J. & KELLY, M. P. 2016b. A systematic literature review of the key challenges for developing the structure of public health economic models. Int J Public Health, 61, 289-98.

STEVENS, W., HILLSDON, M., THOROGOOD, M. & MCARDLE, D. 1998. Cost-effectiveness of a primary care based physical activity intervention in 45-74 year old men and women: a randomised controlled trial. Br J Sports Med, 32, 236-41.

STRONG, M. & OAKLEY, J. E. 2014. When Is a Model Good Enough? Deriving the Expected Value of Model Improvement via Specifying Internal Model Discrepancies. Siam-Asa Journal on Uncertainty Quantification, 2, 106-125.

SULLIVAN, P. W. & GHUSHCHYAN, V. H. 2016. EQ-5D Scores for Diabetes-Related Comorbidities. Value in Health, 19, 1002-1008.

SULLIVAN, P. W., LAWRENCE, W. F. & GHUSHCHYAN, V. 2005. A national catalog of preference-based scores for chronic conditions in the United States. Medical Care, 43, 736-749.

SUTHERLAND, R., REEVES, P., CAMPBELL, E., LUBANS, D. R., MORGAN, P. J., NATHAN, N., WOLFENDEN, L., OKELY, A. D., GILLHAM, K., DAVIES, L. & WIGGERS, J. 2016. Cost effectiveness of a multi-component school-based physical activity intervention targeting adolescents: the 'Physical Activity 4 Everyone' cluster randomized trial. International Journal of Behavioral Nutrition and Physical Activity, 13.

TAGLIAFERRI, F., HAYES, B. P., VIOLA, I. M. & DJOKIC, S. Z. 2016. Wind modelling with nested Markov chains. Journal of Wind Engineering and Industrial Aerodynamics, 157, 118-124.

TAYLOR, D. C. A., PAWAR, V., KRUZIKAS, D., GILMORE, K. E., PANDYA, A., ISKANDAR, R. & WEINSTEIN, M. C. 2010. Methods of Model Calibration Observations from a Mathematical Model of Cervical Cancer. Pharmacoeconomics, 28, 995-1000.

THALER, R. H., & SUNSTEIN, C. R 2008. Nudge: Improving decisions about health, wealth, and happiness, New Haven CT, Yale University Press.

THE BEHAVIOURAL INSIGHTS TEAM 2010. Applying behavioural insight to health. In: OFFICE, T. C. (ed.). London: The Behavioural Insights Team.

THE BEHAVIOURAL INSIGHTS TEAM 2012. EAST Four simple ways to apply behavioural insights. London.

THE CABINET OFFICE 2011. A Guide to Social Return on Investment.

THE KING'S FUND. 2016. Health and wellbeing boards (HWBs) explained [Online]. Available: https://www.kingsfund.org.uk/projects/new-nhs/health-and-wellbeing-boards [Accessed].

THE KING'S FUND, 2018. Spending on Public Health [Online]. Available:

https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/spending-public-health) [Accessed 2 November 2019]

THE SCOTTISH GOVERNMENT 2016. Computable General Equilibrium modelling. In: DIRECTORATE, C. E. (ed.).

THE SROI NETWORK. 2012. A guide to Social return of investment. [Online]. Available: http://www.socialvalueuk.org/app/uploads/2016/03/The%20Guide%20to%20Social%20Return%20on%20Investment%202015.pdf [Accessed 12 November 2019]

THOKALA, P., DEVLIN, N., MARSH, K., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONGRENN, T., MUSSEN, F., PEACOCK, S., WATKINS, J. & IJZERMAN, M. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making-An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value in Health, 19, 1-13.

THOMAS, C. M. & MORRIS, S. 2003. Cost of depression among adults in England in 2000. British Journal of Psychiatry, 183, 514-519.

TOWNSEND P., D. N. 1988. Inequalities in health: the Black Report, London, Pelican Books.

UK GOVERNMENT 2012. Health and Social Care Act 2012.

VAN DYCK, D., MERTENS, L., CARDON, G., DE COCKER, K. & DE BOURDEAUDHUIJ, I. 2017. Opinions Toward Physical Activity, Sedentary Behavior, and Interventions to Stimulate Active Living During Early Retirement: A Qualitative Study in Recently Retired Adults. J Aging Phys Act, 25, 277-286.

VAN KIPPERSLUIS, H. & GALAMA, T. J. 2014. Wealth and Health Behavior: Testing the Concept of a Health Cost. Eur Econ Rev, 72, 197-220.

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.

VEMER, P., CORRO RAMOS, I., VAN VOORN, G. A., AL, M. J. & FEENSTRA, T. L. 2016. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. Pharmacoeconomics, 34, 349-61.

VERGUET, S., KIM, J. J. & JAMISON, D. T. 2016. Extended Cost-Effectiveness Analysis for Health Policy Assessment: A Tutorial. Pharmacoeconomics, 34, 913-923.

VERHOEF, T. I., TREND, V., KELLY, B., ROBINSON, N., FOX, P. & MORRIS, S. 2016. Cost-effectiveness analysis of offering free leisure centre memberships to physically inactive members of the public receiving state benefits: a case study. BMC Public Health, 16, 616.

VESTERGAARD, S., KRONBORG ANDERSEN, C., KORSHOLM, L. & PUGGAARD, L. 2006. Exercise intervention of 65+-year-old men and women: functional ability and health care costs. Aging Clin Exp Res, 18, 227-34.

VLACHOPOULOS, C., AZNAOURIDIS, K., TERENTES-PRINTZIOS, D., IOAKEIMIDIS, N. & STEFANADIS, C. 2012. Prediction of Cardiovascular Events and All-Cause Mortality With Brachial-Ankle Elasticity Index A Systematic Review and Meta-Analysis. Hypertension, 60, 556-+.

VLAEV, I., KING, D., DOLAN, P., and DARZI, A., 2016 The theory and practice of "nudging": changing health behaviors. Public Administration Review, 76 (4). pp. 550- 561. ISSN 0033-3352

VON NEUMANN J., M. O. 1944. Theory of Games and Economic Behavior, Princeton, Princeton University Press.

WAGSTAFF, A. 1986. The demand for health: a simplified Grossman model. . Bull. Econ. Res. 38(1), 93–95 (1986), 93–95.

WALKER, D., AND KUMARANAYAHE, L. 2002. Allowing for differential timing in cost analyses: discounting and annualization, Health Policy and Planning, Volume 17, Issue 1, March 2002, Pages 112–118, https://doi.org/10.1093/heapol/17.1.112

WALTERS, K., RAIT, G., GRIFFIN, M., BUSZEWICZ, M. & NAZARETH, I. 2012. Recent Trends in the Incidence of Anxiety Diagnoses and Symptoms in Primary Care. Plos One, 7.

WANG, G., MACERA, C. A., SCUDDER-SOUCIE, B., SCHIMD, T., PRATT, M. & BUCHNER, D. 2005. A cost-benefit analysis of physical activity using bike/pedestrian trails. Health Promotion Practice, 174-9.

WANG, H., LI, T., SIAHPUSH, M., CHEN, L. W. & HUBERTY, J. 2017. Cost-Effectiveness of Ready for Recess to Promote Physical Activity in Children. Journal of School Health, 87, 278-285.

WANLESS, D. 2003. Securing Good Health for the Whole Population.

WEATHERLY, H., 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.

WHYNES, D., FREW. E., WOLSTENHOLME, J., 2003. Evidence of range bias in contingent valuation payment scales. Health Econ;13:183–90)

WILKINSON, R. G. 1994. The epidemiological transition: from material scarcity to social disadvantage? Daedalus, 123, 61-77.

WOOLDRIDGE, J. 2011. Econometric analysis of cross section and panel data, Cambridge, Mass.: MIT.

WOLFENSTETTER, S. and WENIG, C. 2011. Costing of physical activity programmes in primary prevention: a review of the literature, Health Economics Review, 1, issue 1, p. 1-15, https://EconPapers.repec.org/RePEc:spr:hecrev:v:1:y:2011:i:1:p:1-15:10.1186/2191-1991-1-17.

WORLD HEALTH ORGANIZATION. 1986. First International Conference on Health Promotion. In: ORGANIZATION, W. H., ed. First International Conference on Health Promotion, 1986 Ottawa.

WORLD HEALTH ORGANIZATION. 2006. Body mass index - BMI [Online]. Available: http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi [Accessed].

WORLD HEALTH ORGANIZATION. 2008. Physical activity [Online]. Available: http://www.euro.who.int/en/health-topics/disease-prevention/physical-activity/physical-activity [Accessed 2 November 2019].

WORLD HEALTH ORGANIZATION. 2013. 10 facts on non-communicable diseases [Online]. Available: https://www.who.int/features/factfiles/noncommunicable_diseases/en/ [Accessed].

WORLD HEALTH ORGANIZATION. 2015. Promoting Health, Preventive Disease: The economic case [Online]. Available: http://www.euro.who.int/__data/assets/pdf_file/0006/283695/Promoting-Health-Preventing-Disease-Economic-Case.pdf?ua=1 [Accessed 10 November 2019]

WORLD HEALTH ORGANIZATION. 2016. What is health promotion? [Online]. Available: https://www.who.int/features/qa/health-promotion/en/ [Accessed].

WU, S., COHEN, D., SHI, Y., PEARSON, M. & STURM, R. 2011. Economic analysis of physical activity interventions. American Journal of Preventive Medicine, 40, 149-158.

YANG, Y., SCHUMANN, M., LE, S. & CHENG, S. 2018. Reliability and validity of a new accelerometer-based device for detecting physical activities and energy expenditure. PeerJ, 6, e5775

ZAPATA-DIOMEDI, B., KNIBBS, L. D., WARE, R. S., HEESCH, K. C., TAINIO, M., WOODCOCK, J. & VEERMAN, J. L. 2017. A shift from motorised travel to active transport: What are the potential health gains for an Australian city? PLoS ONE [Electronic Resource], 12, e0184799.

Appendix A

Flow chart of the included studies

```
┌──────────────────────────────────────────────────┐
│ 6,951 potentially relevant records retrieved by    │
│ searching the following electronic databases:      │
│                                                    │
│ • MEDLINE (through OVID; n=1,069)                  │
│ • EMBASE (through OVID; n=3,126)                   │
│ • COCHRANE LIBRARY (n=529)                         │
│ • SportDiscus (through EBSCO; n=1,515)            │
│ • EconLit (through EBSCO; n=712)                  │
└──────────────────────────────────────────────────┘
            │
            │────────────────────►  ┌─────────────────────────────┐
            │                        │ 2,082 duplicates removed     │
            ▼                        └─────────────────────────────┘
┌──────────────────────────────────┐
│ 4,869 records screened by title   │
└──────────────────────────────────┘
            │
            │────────────────────►  ┌─────────────────────────────┐
            │                        │ 4,737 records excluded after │
            ▼                        │ screening titles             │
┌──────────────────────────────────┐ └─────────────────────────────┘
│ 188 records screened by abstract  │
└──────────────────────────────────┘
            │
            │                ┌──────────────────────────────────────┐
            │                │ 134 records excluded after screening  │
            │                │ abstract, of which:                   │
            │───────────────►│                                       │
            │                │ 67 no full EE                         │
            │                │ 48 not solely on PA promotion         │
            ▼                │ 19 not on non-clinical populations    │
┌──────────────────────────────────┐ └──────────────────────────────┘
│ 54 records included for full text │
│ assessment                        │
└──────────────────────────────────┘
            │
            │           ┌──────────────────────────────────────────┐
            │           │ 16 records excluded after screening full  │
            │           │ text, of which:                           │
            │──────────►│                                           │
            │           │ 3 abstracts (full texts not available,    │
            │           │ authors not contactable)                  │
            │           │ 9 no full EE                              │
            │           │ 1 not solely on PA promotion             │
            ▼           │ 3 not on non-clinical populations         │
┌──────────────────────────────────┐ └──────────────────────────────┘
│ 38 EEs included in the systematic │
│ review                            │
└──────────────────────────────────┘
```

Appendix B

Systematic review: data extraction tables

B.1 Overview of the review studies

Reference	Country	Promotion level	Targeted determinants of health / population	Promotion setting	Policy category	Intervention approach / components	Cost per outcome unit
Amarasinghe 2010	Australia	Targeted	PI	Primary care	Communication	GP advice	AU $ 11,000 / DALY gained
Anoyke et al., 2013	UK	Targeted	PI	Primary care	Communication	Brief advice	UK £1,730/QALY gained
Babey et al., 2014	USA	Universal	general population (pupils)	Occupational (school)	MULTIPLE - Service provision, Communication/ Marketing, Environmental	4 types of school-based opportunities: before, after school, augmented PE and short PA breaks	in-class US $ >0.01/MET-hour gained; before-school US $0.49/MET-hour gained; longer day US $ 0.65-0.98/MET-hour gained; after-school US $ 10.62/MET-hour gained

Barrett et al., 2015	USA	Universal	general population (pupils)	Occupational (school)	MULTIPLE - Service provision, Communication/ Marketing, Environmental	Active physical education policy	US $ 0.34/MET-hour-day gained
Beale et al., 2012	UK	Universal	general population	Community-based	Environmental	Built environment	UK £100 - 10,000/QALY gained; £11 per £1 invested
Cavill, 2011	UK	Universal	general population (cyclists)	Community-based	MULTIPLE - Environmental + Marketing	Multicomponent programme: built environment, promotion and smart measures	US $2.6-3.5 per $1 invested
Chen et al. 2008	Taiwan	Universal	general population	Community-based	Service provision	Hospital-based, 12 week, supervised walking programme	US $21,936/QALY gained - (US thresholds $50,000=acceptable; $20,000 definitively acceptable)
Cobiac et al. 2009	Australia	Combination	general population	Combination	MULTIPLE - Communication/ Marketing, Service provision, Environmental	Multiple interventions / GP prescription, referral, campaigns, active transport, pedometers, internet	Pedometers=dominant; mass-media campaign= dominant; internet-based program AU $3,000/DALY gained; GP prescription AU $12,000/DALY gained; active transport program AU $20,000/DALY gained; GP referral AU $79,000/DALY gained (AU $ 50,000/DALY gained threshold

Dallat et al., 2013	UK (Northern Ireland)	Universal	general population	Community-based	Environmental	Urban regeneration project	UK £4,469/DALY gained
De Schmedt et al. 2011	Belgium	Universal	general population	Community-based	MULTIPLE - Environmental, Communication/ Marketing, Service provision	Multicomponent programme=media campaign, environmental approaches, the sale and loan of pedometers, and several local physical activity projects vs no programme	DOMINANT (KCE recommended threshold of EUR 30,000/QALY)
Frew et al. 2014	UK (England)	Universal	general population	Community-based	Service provision	Programme based on free access to public leisure centres, located in deprived city areas, during off-peak times	UK £400/QALY gained; + £96 net benefit value of programme

Golsteijn et al. 2014	The Netherlands	Universal	general population	Community-based	Marketing	Four modalities of printed / web-based information	(average of 4 interventions) EUR -46/MET-hours-week; EUR 101,169/QALY gained (Web-based environmental EUR -47,293/QALY gained) - (WTP=EUR 20,000/QALY gained)
Goyder et al. 2014	UK (England)	Targeted	PI, age, economic deprivation	Primary care	Communication	(motivational interview-based) face-to-face vs telephone based counselling	no aggregate results
Groessl et al., 2015	USA	Targeted	PI, age, increased CD risk	Community-based	Service provision	2 programmes: times/week, centre-based, exercise programme + home-based activity vs a weekly (first 26-week and then monthly) health education programme	US $42,376/MMD prevented; US $49,167/QALY gained
Gulliford et al. 2014	UK	Targeted	PI	Primary care	Communication	GP advice	3.2 QALY per 1,000 participants (valuing £30,000 one QALY)
Guo and Gandavarapu 2010	USA (Winsconsin)	Universal	general population	Community-based	Environmental	Built environment (adding sidewalks)	US $1.87 per $1 invested

Gusi et al., 2008	Spain	Targeted	Age, weight status, increased CD risk	Primary care	Service provision	A 6-month, 3 times/week, walking-based, supervised exercise programme vs "best practice"	EUR 311/QALY gained
Haas M., 2006	New Zealand	Universal	Elderly (>60)	Community-based	Service provision	A 1 time/week, centre-based, discounted (free first 5 sessions) Tai Chi programme vs 1 control= waiting-list	no aggregate results
Isaacs AJ et al., 2007	UK (England)	Targeted	PI, age, increased CD risk	Primary care	MULTIPLE - Service provision, Communication	2 10-week, 2-3 times/week intervention conditions: supervised exercise classes in local leisure centre; instructor-led walking programme.	no aggregate results
Larsen et al.2015	USA (Rhode Island)	Targeted	PI, ethnicity, gender	Community-based	Communication	Individually tailored PA counselling vs mail-based advice	US $ 0.15/MVPA minute gained at 6 months; 0.05 at 12 months
McEachan et al., 2011	UK (England)	Universal	general population (employees)	Occupational (workplace)	Communication	PA facilitators	(-) £103.02 (negative net benefit valuing £20,000 one QALY)

Montes et al. 2011	America	Universal	general population (users)	Community-based	MULTIPLE - Environmental, Communication/ Marketing	Multicomponent programme: environment (closed streets) + promotion	US $1.02 - 4.26 per $1 invested
Moodie et al., 2009	Australia	Universal	general population (elementary pupils)	Occupational (school)	Service provision	Walking school bus with 2 conductors vs current practice (do-nothing)	AU $ 760,000 / DALY gained
Munro et al. 1997	UK (England)	Universal	general population	Community-based	Service provision	Free, twice a week, community-based, exercise programme	£ 330/LYS
Murphy et al., 2012	UK (Wales)	Targeted	PI, increased CD risk	Primary care	Service provision	Exercise referral scheme (motivational interview-based)	£ 12,111/QALY gained
Nshimyumuk iza et al., 2012	Canada	Universal	Women >40 from general population	not specified	Marketing	Mass-media campaign	dominant (CAD $ 50,000 / fracture averted or QALY gained)
Over et al., 2012	The Netherlands	Targeted	PI	Primary care	Communication	GP counselling + pedometer	EUR 13,200/LYS; EUR 11,100/QALY gained
Munro 2004	UK (England)	Targeted	PI	Community-based	Service provision	A 24-month, 2 times/week programmes in 4 GPs=community centres, free exercise classes	EUR 17,174/QALY gained
Peterson et al. 2008	USA (Delaware)	Universal	general population	Community-based	Marketing	Mass-media campaign	US $ 8.87 per person becoming more active

Pringle et al. 2010	UK	Combination	general population in high-need UK areas	Combination	MULTIPLE - Communication/ Marketing, Service provision, Environmental	Multiple interventions: media campaigns, exercise classes, referrals, motivational interviews and outdoor activities	UK £ 260 - 2,786 per person becoming more active; £ 47 - 509/QALY gained
Roux et al. 2008	USA	Combination	general population	Community-based	MULTIPLE - Communication/ Marketing, Service provision, Environmental	Multiple interventions: community-wide campaigns, tailored behaviour change, social support and enhanced access to places for PA	US $ 22,654 - 110,322/LYS; $ 14,286-68,557/QALY gained (willingness-to-pay threshold at $ 200,000 per QALY)
Roux et al. 2015	USA	Combination	general population	Community-based	MULTIPLE - Communication/ Marketing, Service provision, Environmental	Multiple interventions: community-wide campaigns, tailored behaviour change, social support and enhanced access to places for PA	US $ 63,737 - 237,933/LYS; $ 33,639 - 127,464/QALY gained (willingness-to-pay threshold at $ 200,000 per QALY)
Sevick et al. 2000	USA (Texas)	Targeted	PI, age, weight status	Community-based	Communication vs service provision	Centre-based counselling vs structured exercise programme	Lifestyle: US $ 2/walking minute gained; structured US $ 7/walking minute gained (at 6 months). $ 1 each at 12 months

Shaw et al, 2011	UK (Scotland)	Targeted	PI, economic deprivation	Community-based	MULTIPLE - Communication/ Marketing, Service provision	2 36-week walking intervention conditions initial 12-week waiting-list period for "minimal" intervention; "maximal"="minimal" + 30 mins consultation	minimal £92; maximal £591/per person achieving PA target
Stevens et al. 1998	UK (England)	Targeted	PI	Community-based	Service provision	A personalised 10 week programme vs home based activities	UK £650/per person moving out of the sedentary group
Sutherland et al., 2016	Australia	Universal	Secondary schools in disadvantaged communities	Occupational (school)	MULTIPLE - Service provision, Communication/ Marketing, Environmental	Multicomponent school-based programme 24-month intervention incorporated as part of usual school business	AU $56/MVPA minute gained; $1/MET-hour gained; $1408/unit of BMI avoided
Vestergaard et al., 2006	Denmark	Universal	general population	Community-based	Service provision	A 1 time/week, training facility-based, exercise programme	no aggregate results
Wang et al. 2005	USA (Nebraska)	Universal	general population	Community-based	Environmental	Built environment (bike/pedestrian trails)	US $2.94 per $1 invested

CD=chronic disease; KCE= (Belgian) health care knowledge centre; LYS=life years saved; PE=physical education; PI=physical inactivity.

B.2   Analytical methods used for estimation of effectiveness

Reference	Analysis year/s	Effectiveness study design (setting) / max follow-up period	Estimation method	Heterogeneity of effect	Selection bias
Cavill 2011	2005-2009	Pragmatic (no control, community based) programme / 3 years	before/after approach, surveys	not reported	not reported
Chen et al. 2008	2007	prospective controlled trial (community-based) 1:1 / 12 weeks	controlled study	age, gender and other socio demographics	not reported
De Schmedt et al. 2011	2005-2006	Cluster (2 small cities) controlled (community-based) programme - 1 year	controlled study	age, gender and other socio demographics	not reported
Frew et al. 2014	2010	Pragmatic (no control, community-based) programme / 3-4 months	before/after approach, surveys	not reported	not reported
Golsteijn et al. 2014	2011	Cluster (6 municipal health regions, MHR - 14 neighbourhoods, community-based) RCT 1:2:1:1:1 / 1 year	controlled study	age, gender and other socio demographics	LOCF + linear interpolation methods for the 6 to 12 months
Guo and Gandavarapu 2010	2001	Observational evidence (travel survey data) / estimated (equation) shift in walking and cycling levels	SSUR model, econometric aggregate level survey data	aggregate level analysis	reported to have accounted for

Haas M., 2006	2002	RCT (community-based) 1:1 / 16 weeks	controlled study	not reported	not reported
McEachan et al., 2011	2007-2008	Cluster (5 organisations, 44 worksites) RCT 1:1 / 1 year	controlled study, multilevel modelling	age, gender and socio-economic status	not reported
Montes et al. 2011	2005 - 2010	4 pragmatic (no control, community based) programmes / assumed 1 year	before/after approach, surveys	not reported	not reported
Peterson et al. 2008	2004	pragmatic (no control) programme / 6 weeks	before/after approach, surveys	not reported	not reported
Pringle et al. 2010	2004 - 2006	Several controlled studies - 7 (community-based) intervention modalities / minimum 6 weeks	controlled studies	not reported	not reported
Sutherland et al., 2016	2012-2014	Cluster (10 secondary schools in disadvantaged communities) RCT 1:1 / 2 years	controlled study	by cluster	not reported
Vestergaard et al., 2006	2002	pragmatic (no control) programme (community-based) / 5 months	before/after approach, surveys	age and gender	not reported
Wang et al. 2005	1998	Observational evidence - census reports / assumed shift in n. of trail uses	cross-sectional surveys	not reported	not reported

LOCF=last observation carried forward

B.3   Modelling methods used to extrapolate effects over time

Reference	Analysis sample	Target population	Evaluation time horizon	Individual / aggregate level	Timed / untime modelling	Modelling paradigm	Range of diseases	Time lag to health benefit	Decay of effects over time
Amarasinghe 2010	Prevalence data	Western Australia	lifetime	Aggregate level	Untimed	CRA	T2D, HD,STR,CC,DEP	not specified	constant, different rates of compliance
Anoyke et al., 2013	100,000 healthy inactive adults aged 33	Adults	lifetime	Aggregate level	Timed	Markov chains	T2D,CHD,STR	1 year run-in period	100% constant, benefits accrue for the first 10 years over lifetime
Barrett et al., 2015	2015 US population, no other details specified	US school pupils aged 6-11 y, N=18.5 million	10 y	Aggregate level	Timed	Markov chains	obesity	2 years to full effect on BMI	100% constant, benefits accrue over the whole TH
Beale et al., 2012	1,000 sedentary adults	Adults	10 y	Aggregate level	Timed	Markov chains	T2D,CHD,STR	not specified	100% constant, benefits accrue over the whole TH
Beale et al., 2012	1,000 sedentary adults	Adults	30 y	Aggregate level	Untimed	CRA	not explicit	not specified	100% constant, benefits accrue over the whole TH
Cavill, 2010	N=not specified, 16 y+	not specified	10 y	Aggregate level	Untimed	Off-the-shelf tools, HEAT (CRA) & WebTAG	not explicit	5 years	100% constant, benefits accrue over the whole TH

Cobiac et al. 2009	Population, age/gender distribution	Australia population	lifetime	Aggregate level	Timed	Multiple cohort life-table approach	T2D, CHD,IHD, STR, BRC	not specified	Sustained for the first year, but decay exponentially at a rate of 50% per annum thereafter
Dallat et al., 2013	Prevalence data	N=110,600, 16 y+	41 y	Aggregate level	Untimed	Off-the-shelf tool, PREVENT (CRA)	T2D, IHD, STR,CC,BRC	not specified for base-case analysis, 2 to 20 years in SA	100% constant, benefits accrue over the whole TH
De Schmedt et al. 2011	N=266 adults 25-75 y who improved PA level	N=245,000, adults 25-75 y	20 y	Aggregate level	Timed	Markov chains	T2D,CHD,STR, CC	not specified	100% constant, benefits accrue over the whole TH
Frew et al. 2014	N=not specified, population of adults 16-70 y, no other details provided	City adult population	lifetime	Aggregate level	Timed	Markov chains	T2D, CHD, STR, CRC, BRC	not specified	100% constant, benefits accrue over the whole TH, 50% after the first year in SA
Goyder et al. 2014	N=500,000 age/gender matched individuals	Sedentary adults from deprived areas	lifetime	Individual level	Timed	Markov chains	not explicit	not specified	Decline of 100% after the 2 years, benefits accrue over the whole TH

Gulliford et al. 2014	N=262,704 inactive adults> 30 y	UK population	5 and 10 y	Aggregate level	Timed	Markov chains	T2D, CHD, STR, CRC + DEP interaction (32 combinations)	not specified	100% constant, benefits accrue over the whole TH
Guo and Gandavarapu 2010	N=438,881	County population	10 y	Individual level	Untimed	System of linear equations	obesity	not specified	100% constant, benefits accrue over the whole TH
Montes et al. 2011	N=not specified, 16 y+	National population	5 and 10 y	Aggregate level	Untimed	Off-the-shelf tools, "HEAT" (CRA)	not explicit	not specified	100% constant, benefits accrue over the whole TH
Moodie et al., 2009	N=15,680 average school pupils 5-7 y	Australia primary school children	lifetime	Aggregate level	Untimed	CRA	obesity	not specified	100% constant, benefits accrue over the whole TH
Munro et al. 1997	N=10,000 adults>=65 y	Older adults	10 y	Aggregate level	Untimed	CRA	not explicit	not specified	100% constant, benefits accrue over the whole TH
Nshimyumukiza et al., 2012	N=500,000, women>= 40 y	Women>=40y	5 and 10 y	Individual level	Timed	Individual-level Markov model	Osteoporosis	not specified	100% constant, benefits accrue over the whole TH

Over et al., 2012	N=not specified, population 20-65 y	National population 20-65 y	lifetime	Aggregate level	Timed	Markov chains, existing model	not explicit	not specified	explicit constant at 25% after 18 weeks, benefits accrue over the whole TH
Pringle et al. 2010	N=not specified, population 10+ y	Adults	lifetime	Aggregate level	Untimed	Decision-tree, existing model	T2D, CHD,STR,CC	not specified	50% constant, benefits accrue over the whole TH
Roux et al. 2008	N=not specified,adult population 25-64 years	Aduls 25-64 years	lifetime	Aggregate level	Timed	Markov chains	T2D,CHD,STR, CRC,BRC	not specified	explicit 33% to 50% of decline in effect after year 2, benefits accrue over the whole TH
Roux et al. 2015	N=not specified,adult population 50-64 years	Adults 50-64 years	lifetime	Aggregate level	Timed	Markov chains	T2D,CHD,STR, CRC,BRC	not specified	explicit 33% to 50% of decline in effect after year 2, benefits accrue over the whole TH
Wang et al. 2005	N=not specified, trail users	National population	30 years	Aggregate level	Untimed	CRA	not explicit	not specified	100% constant, benefits accrue over the whole TH

BRC=Breast Cancer, CC=colon Cancer, CRC=Colorectal Cancer, CHD=Coronary Heart Disease, CRA=Comparative Risk Assessment, STR=Stroke; T2D=Type II Diabetes, TH=Time Horizon, y=years

B.4 Methods used for economic analysis – part 1

Reference	Form/s of economic evaluation (measure of benefit)	Stated perspective/s	Comparator	Equity considerations	Spill-over effects	Non-health effects	Uncertainty (details)
Babey et al., 2014	CEA (MET)	not specified	(implicit) no intervention	Discussed implementation and acceptability issues - differences in availability of space	none	none	none
Barrett et al., 2015	CEA (MET) + CCA	modified societal perspective	current practice	Discussed implementation issues potentially increasing inequities	Argued that no compensatory effects + intervention able to change social norm + trained teachers may be effective in promoting movement in other parts of the school day	none	one-way SA (intervention costs, effect size, implementation rate), PSA
Cavill, 2011	CBA (cost-benefit ratios)	not specified (public investor?)	(implicit) no intervention	none	Argued that no compensatory effects would take place	travel-related costs and benefits including decongestion, absenteeism, amenity, accidents	one-way SA (time horizon)

Chen et al. 2008	CUA (QALY) + CCA	not specified (health care sector?)	no intervention	none	none	none	none
Cobiac et al. 2009	CUA (DALY) + CCA	health sector	current practice	none	Mentioned possible synergistic effects with implementation of multiple interventions	none	one-way SA (dissipation of effect size)
De Schmedt et al. 2011	CUA (QALY)	public payer (health sector?)	no intervention	none	none	none	one-way SA (RR, healthcare and intervention costs, "utilities"), PSA
Frew et al. 2014	CUA (QALY) and CBA (WTP exercise)	healthcare (CEA + "wider" (CBA)	no intervention	Acknowledged: Sub-group analysis is limited because of lack of power	none	none	one-way SA (dissipation of intervention effect on part of the sample - time horizon, start-up costs)

Golsteijn et al. 2014	CEA (MET) and CUA (QALY)	societal (+ healthcare in SA)	(waiting-list) no intervention	Acknowledged: Implementation in inactive population	none	participant and family costs, travel costs, productivity losses	one-way SA (baseline differences between the 4 conditions), PSA
Guo and Gandavarapu 2010	CBA (cost/benefit ratios)	not specified (public investor?)	(implicit) no intervention	none	Mentioned: estimated substantive, synergistic and complementary effects of built environments changes	air quality benefits	none
Haas M., 2006	CCA	health system	(implicit) no intervention	none	none	none	one-way SA (intervention cost)
McEachan et al., 2011	CUA (net benefit in QALY) + CCA	societal	(waiting-list) no intervention	none	none	time cost	PSA

Montes et al. 2011	CBA (cost/benefit ratios)	public health	no intervention	none	none	none	one-way SA (healthcare costs, n. of users), PSA
Moodie et al., 2009	CUA (DALY)	societal	(implicit) no intervention	Discussed implementation and acceptability issues	Discussed: "spin-offs" to both the wider student population as well as to parents and the wider community- - "second stage filter analysis"	cost to participants (no details)	one-way SA (intervention effect, costs), PSA, Scenario analysis (costs attribution)
Munro et al. 1997	CEA (LYS) + CCA	healthcare provider	(implicit) no intervention	none	none	none	one-way SA (effect size, intervention cost, incidence reduction, life expectancy, adherence)
Peterson et al. 2008	CEA (per person becoming more active)	not specified (public investor?)	(implicit) no intervention	none	none	not clear	none

Pringle et al. 2010	CEA (per person becoming more active) + CUA (QALY) + CCA	healthcare	(implicit) no intervention	none	none	none	none
Roux et al. 2008	CEA (LYS) + CUA (QALY)	societal	no intervention	Acknowledged: Sub-group analysis is limited because of limited data	none	time and productivity costs	one-way SA (dissipation of effect size and time horizon), PSA
Roux et al. 2015	CEA (LYS) + CUA (QALY)	societal	no intervention	none	none	time and productivity costs	PSA (intervention effect size and costs)
Sutherland et al., 2016	CEA (MVPA minute, MET, BMI units)	societal	current practice	none	none	none	one way SA (intervention cost, effect size), Scenario analyses (dissemination)

Vestergaard et al., 2006	CCA	health care sector	(implicit) no intervention	Discussion about barriers given by transport - accessibility issues for potential participants)	none	none	one-way SA (unit costs)
Wang et al. 2005	CBA (cost-benefit ratios)	public health	(implicit) no intervention	none	none	none	one-way SA (intervention costs)

CBA=cost-benefit analysis, CCA=cost-consequences analysis, CEA=cost-effectiveness analysis, CUA=cost-utility analysis, MET=metabolic equivalent of task, MVPA=moderate to vigorous physical activity, PSA=probabilistic sensitivity analysis, SA=sensitivity analysis, WTP=willingness to pay.

B.5   Methods used for economic analysis – part 2

Reference	Start-up costs	Delivery/ running costs	Intervention components / cost drivers	Main assumptions and costing rules	Non-health costs	Health-care costs	Health-related consequences
Amarasinghe 2010	no	yes	subsidy for GP advice	none	none	(direct) cost for disease treatment	DALY
Anoyke et al., 2013	not relevant	yes	consultation with GP	none	none	(direct) cost for disease treatment	morbidity and mortality prevention - QALY
Babey et al., 2014	no	yes	Personnel, equipment, supplies and material, overhead costs and transport	Programme operating costs only, overhead costs	costs by families for enrolment in the program	none	MVPA, MET
Barrett et al., 2015	yes (no research and development)	yes	Personnel, equipment, PE curricula.	As if operating under steady-state conditions	additional training time for facilitators	health care cost savings from obesity prevention	MVPA, MET
Beale et al., 2012	yes	yes	construction and maintenance	Assumed 30 years life cycle - evenly allocated	none	(direct) cost for disease treatment	morbidity

Cavill, 2011	not specified	not specified	budget expenditure	Assumed that costs are incurring in the first 3 years but benefits over 10 years	none	Value of statistical life: net benefit from reduced mortality and dis-benefit from n. of accidents	mortality prevention (no morbidity), n. of accidents - medical cost savings
Chen et al. 2008	no	yes	Personnel, administration, equipment maintenance	none	transportation fees, lost income from informal care, cost of extra-equipment and baby-sitting.	none	health care utilisation: n. of hospitalisations, outpatient and emergency visits
Cobiac et al. 2009	no	yes	6 intervention types - according to the respective studies	As if operating under steady-state conditions	time (not specified) and travel costs for participants	(direct) cost for disease treatment	(n. of people moving PA level) cases averted, DALY, mortality
Dallat et al., 2013	yes	yes	construction and maintenance costs	Assumed 41 years life cycle - evenly allocated	none	(direct) cost for disease treatment	morbidity, mortality, medical costs savings, DALY
De Schmedt et al. 2011	yes	yes	Promotion materials, development and maintenance of website, 1/2 full time equivalent staff and pedometers for users only	Assumed 5 years life cycle for pedometers	none	(direct) cost for disease treatment + diabetes complications	QALY

Frew et al. 2014	yes (but with facilities already existing and running)	yes	Income replacement (89%), gym refurbishment, marketing, monitoring, technical support, leisure card, extended offer, project management.	changing annual usage rate of 50%-100% to account for the effects of changing levels of participation	none	(direct) cost for disease treatment	QALY
Golsteijn et al. 2014	yes	yes	invitations, printing and postage, staffing costs for handling questionnaires, advice and reminders, gathering environmental info and hosting costs for tailoring software and website (no research)	not specified; friction costs method for productivity loss	out-of-pocket by participants: family and personal (sport membership, equipment) and travel + productivity loss + exercise time not valued as a cost as assumed increased QoL in leisure time	For health care (e.g. nights in hospital, lifestyle coach, medical specialist)	(METs and) QALY
Goyder et al. 2014	no	yes	Personnel, training, venue hire (for the community-based interviews), phone	none	Use of NHS facility time valuation: (value of time - average wage)	none	morbility and mortality rates incorporated in QALY

Groessl et al, 2015	no	yes	Direct costs: material, incentives (gift cards), refreshments, personnel (time) + overheads (estimated 69% personnel costs for facilities costs, indirect support personnel and other indirect costs)	Only delivery costs - overheads f(x) personnel costs	none	Health care utilization	Major mobility disability (walking test), quality of well-being (to produce QALY)
Gulliford et al. 2014	not relevant	yes	consultation with GP	active participants 20% of intervention cost	none	Health care utilization (primary, secondary care and prescriptions)	morbidity prevention - QALY
Guo and Gandavarapu 2010	yes	no	construction costs only	Assumed 10 years life cycle	none	(direct) medical cost associated with obesity	Obesity-related outcomes
Gusi et al., 2008	no	yes	salary of a graduate sport sciences only	no marginal societal costs, recruitment did not require additional time	none	health care utilization (medications, consultations)	BMI, anxiety and depression + QALY
Haas et al. 2006	no	yes	venue hire, staff, advertising	Programme operating costs only, overhead costs not considered	none	health care (e.g. hospitalisation - standard costs)	Falls

Isaacs AJ et al., 2007	yes (no research and development)	yes	To the public sector: providing facilities, exercise trainers and administrative support, initial assessment.	data collected during the trial	out-of-pocket by participants: time costs, travel costs, childcare and equipment.	costs averted for use of pharmaceuticals prescribed by GPs and hospital admissions	QALY
Larsen et.al	no	yes	Personnel, overhead, costs of expert system, hardware/software, materials, printing, postage (no recruitment or research)	Straight-line 5 years depreciation method (evenly allocated) for hardware/software: assumed 3 years of use - overhead costs assumed 10%	none	none	(MVPA)
McEachan et al., 2011	yes (development included)	yes	Labour time, equipment, consumables, travel, graphic design (website)	data collected during the trial	out-of-pocket by participants, exercise time and travel costs + impact on productivity due to absence due to ill-health	net monetary benefit as the difference between value of QALY (£20,000) - intervention + other costs ("out-of-pocket")	MVPA changes, QALY
Montes et al. 2011	yes	yes	construction and maintenance costs (excepet for one program with existing infrastructure)	Assumed 10 years life cycle for equipment	costs to the potential exerciser: equipment	Value of statistical life: net benefit from reduced mortality	(transport-related) mortality prevention

Moodie et al., 2009	no	yes	Personnel, volunteers training , insurance + central and local admin	As if operating under steady-state conditions, Central admin, overhead costs and annuitization of fixed costs	opportunity costs (time) by volunteers (25% hourly wage)	health care cost savings from obesity prevention (prevalence-based data?)	DALY
Munro et al. 1997	no	yes	Hire of halls, personnel, refreshments, ongoing publicity and recruitment	As if operating under steady-state conditions	none	(direct) cost for hospitalisation	n. of cases (in-patient), mortality
Murphy et al., 2012	not specified	not specified	intervention cost per participants fixed with no detail disclosure	none	none	none	QALY
Nshimyumukiza et al., 2012	not specified	not specified	prevention campaigns	none	none	(direct) costs for fracture treatment	fracture events, QALY
Over et al., 2012	no	yes	PA checks, counselling, pedometer, follow-up sessions by GP assistant	none	none	none	(morbidity, mortality) QALY
Munro 2004	yes	yes	Recruitment, administration, hire of halls, exercise leaders and refreshments	start-up costs annuitized over 5 years (evenly allocated)	none	none	(habitual PA) health status, mortality, QALY

Peterson et al. 2008	yes	yes	production and placement costs (no description)	not specified	none	none	none
Pringle et al. 2010	yes	yes	Personnel, training, premises, transport, equipment, publicity and other running costs	not specified	none	(direct) cost for disease treatment	(n. of people moving PA level) cases averted, QALY
Roux et al. 2008	no (except for enhanced access - development and maintaining infrastructure)	yes	7 intervention types - according to the respective studies	Time valuation: time for exercising = wage value	out-of-pocket by participants: equipment + exercise time valuation (value of exercise time age-gender specific wage)	(direct) cost for disease treatment	QALY, mortality
Roux et al. 2015	no (except for enhanced access - development and maintaining infrastructure)	yes	7 intervention types - according to the respective studies	Time valuation: time for exercising = wage value	out-of-pocket by participants: equipment + exercise time valuation (value of exercise time age-gender specific wage)	(direct) cost for disease treatment	QALY, mortality

			Personnel, computerised tracking system, curriculum materials, printing and postage, facilities and health club membership fees	personnel cost allocation + estimation of facility maintenance, utilities and telephone costs - method not specified			
Sevick et al. 2000	yes	yes	Personnel, computerised tracking system, curriculum materials, printing and postage, facilities and health club membership fees	personnel cost allocation + estimation of facility maintenance, utilities and telephone costs - method not specified	none	none	habitual PA level, energy expenditure, cardiorespiratory fitness
Shaw et al, 2011	no	yes	Pedometer, follow-up calls and consultation	not specified	none	none	n. of people becoming active
Stevens et al. 1998	yes	yes	questionnaire design and production, mailing and follow up of non-respondents, postage, stationary, labour, equipment (no research)	1/3 of costs for data processing	none	none	(n. of people moving PA level)
Sutherland et al., 2016	yes (no research and development)	yes	Personnel, equipment, materials, printing	data collected during the trial	opportunity costs (time) for activities outside PE	none	MVPA, MET

Vestergaard et al., 2006	not specified	not specified	not specified	unit costs - national level	none	health care and PH services use (hospitalisation, out-patient treatment, GP visits, physiotherapists, other medical specialists)	Mobility, BMI, functional ability
Wang et al. 2005	yes	yes	construction and maintenance costs	assumed 30 years usage (life cycle) - evenly allocated	out-of-pocket by participants: equipment, assumed trail use during leisure time, thus exercise time not valued as a cost	direct medical cost savings	(n. of people meeting PA recommendation)

Notes: see B.4

Appendix C

Leeds Let's Get Active: promotional material

Appendix D

Leeds Let's Get Active questionnaires

D.1   Modified IPAQ questionnaire

Draft Report – August 2015

Appendix A: Example Questionnaire

1. In the past week, on how many days have you done a total of 30 minutes or more of physical activity, which was enough to raise your breathing rate? This may include sport, exercise, and brisk walking or cycling for recreation or to get to and from places, but should not include housework or physical activity that may be part of your job?

    ___   **Days per week**

We want to find out about the kinds of physical activities that people do in everyday life. The questions are about the time you spent being physically active in the <u>last 7 days</u>. Please answer each question even if you think of yourself as inactive. Please think about the activities you do at work, as part of your work around the house or garden, to travel from place to place, and in your spare time for fun, exercise or sport.

*Think about all the **vigorous** things you did in the **last 7 days**. **Vigorous** physical activities take hard physical effort; they make you breathe much harder than normal. Think only about those physical activities that you did for at least 10 minutes at a time.*

2. During the **last 7 days**, on how many days did you do **vigorous** physical activities like heavy lifting, digging, aerobics, or fast bicycling?

    ___ **Days per week**   **OR** [ ] Nothing vigorous ➡ *Skip to Q3*

3. How much time did you usually spend doing **vigorous** physical activities on one of those days?

    _____ **Hours per day**   _____ **minutes per day**   **OR** [ ] ?/Not sure

Think about all the **moderate** activities that you did in the **last 7 days**. **Moderate** activities that take moderate physical effort and that make you breathe a bit harder than normal. Think only about activities that lasted at least 10 minutes at a time.

4. During the **last 7 days**, on how many days did you do **moderate** physical activities like carrying light loads or bicycling at a regular pace? Do not include walking.

    ___ **Days per week**   [ ] none ➡ *Skip to Q5*

33

5. How much time did you usually spend doing **moderate** physical activities on one of those days?

____ **Hours per day**   ____ **minutes per day**   **OR** ☐ ?/Not sure

Think about the time you spent **walking** in the **last 7 days**. This includes at work and at home, walking as travel, and any other walking that you might do.

6. During the **last 7 days**, on how many days did you **walk** for at least 10 minutes at a time?

____ **Days per week** ➡ ☐ No walking   *Skip to Q7*

7. How much time did you usually spend **walking** on one of those days?

____ **Hours per day**   ____ **minutes per day**   **OR** ☐ ?/Not sure

The last question is about the time you spent **sitting** on weekdays during the **last 7 days**. Include time spent at work, at home, while doing course work and during leisure time. This may include time spent sitting at a desk, visiting friends, reading, or sitting or lying down to watch television.

8. During the **last 7 days**, how much time did you spend **sitting** on a **week day**?

____ **Hours per day**   ____ **minutes per day**   **OR** ☐ ?/Not sure

Finally, I'd like you to think about any **Sport** that you have done in the **last 7 days.** By **Sport** we mean any competitive or non-competitive sporting activity. This can include planned exercise sessions such as running or jogging. Think only about those sports or exercises that you did for at least 10 minutes at a time.

9. During the last 7 days, on how many days did you take part in any **sport**?

_____ **Days per week**   ☐ No sport ➡ *Skip to end*

10. How much time did you usually spend doing sport on one of those days?

____ **Hours per day**   ____ **minutes per day**   **OR** ☐ ?/Not sure

This is the end of the questionnaire, thank you for participating.

34

D.2   Lifestyle questionnaire

About You

Person number:
(Staff must complete)

Leeds let's get active!

1   Name _____ Date of Birth ___/___/_____ Age _____

2   Address _____

     Postcode_____ Telephone number_____

     Email Address _____

3   Male ☐   Female ☐

4   Which of the following best describes your ethnic background?

White:	British	Asian or Asian British :	Indian	
	Irish		Pakistani	
			Kashmiri	
Mixed:	White & Black Caribbean		Bangladeshi	
	White & Black African	Black or Black British:	Caribbean	
	White & Asian		African	
Other:	Chinese	Other:	EU National	
	Gypsy/ Traveller		Other (please specify)	

5   What is your current employment status ?

     Full-time ☐   Part-time ☐   Unemployed ☐ Not working due to ill health/disability ☐
     Student ☐   Volunteer ☐   Retired   ☐

6   What is your highest level of academic qualification?

     No Qualifications   GCSE/O Level grade A*- C   A Levels
            ☐   ☐   ☐

     Diploma in Higher   First Degree (e.g. BSc, BA)   Higher Degree (e.g. MSc, PhD)
     Education   ☐   ☐
            ☐

7   Which of the following best describes your relationship status?

     Single   ☐   Married   ☐   Live with partner ☐
     Widowed ☐   Divorced/Separated ☐   Prefer not to say ☐

8   Which of the following best describes how you found out about this project?

     Friends   Word of Mouth   Health professional   Newspaper   Internet   Social media
       ☐   ☐   ☐   ☐   ☐   ☐
     Community Venue   Leisure Centre   Radio   Email   Website_____Other_____
       ☐   ☐   ☐   ☐

B. ABOUT YOUR PHYSICAL ACTIVITY

9   In the past week, on how many days have you done a total of 30 minutes or more of physical activity, which was enough to raise your breathing rate? This may include sport, exercise, and brisk walking or cycling for recreation or to get to and from places, but should not include housework or physical activity that may be part of your job ?

     None   One   Two   Three   Four   Five   Six   Seven
       ☐   ☐   ☐   ☐   ☐   ☐   ☐   ☐

C. ABOUT YOUR LIFESTYLE

10 How many portions of fruit and/or vegetables (including fruit juices and fresh, frozen, dried, juiced and canned fruit, pulses and salad) did you eat yesterday?

None	One	Two	Three	Four	Five	Six	Seven+
❑	❑	❑	❑	❑	❑	❑	❑

11 Do you currently smoke cigarettes, cigars or a pipe?
Never Smoked ❑
Former Smoker ❑
Current Smoker ❑

12 If you 'smoke', how many cigarettes per day on average do you 'smoke'? [   ]

13 How often do you have a drink containing alcohol?

Never	Monthly or less	2-4 times a MONTH	2-3 times a WEEK,	4 or more times a week
❑	❑	❑	❑	❑

14 How many units of alcohol do you drink on a typical day when you are drinking?

1or 2 units	3 or 4 units	5 or 6 units	7, 8 or 9 units	10 or more units
❑	❑	❑	❑	❑

15 How often have you had 6 or more units if female/8 or more if male, on a single occasion in the last year?

Never	Less than monthly	Monthly	Weekly	Daily or almost daily
❑	❑	❑	❑	❑

D. ABOUT YOUR HEALTH

16 In the last 12 months, have you been diagnosed by a GP/Doctor/Health Professional as having any of the following long term conditions/health problems?

Asthma	Cancer	Chronic Kidney Disease	COPD
❑	❑	❑	❑
Coronary Heart Disease ❑	Dementia ❑	Depression ❑	Diabetes ❑
Epilepsy ❑	Hypertension ❑	Mental Health Problem ❑	Stroke ❑

17 Wellbeing

Overall, how satisfied are you with your life nowadays? [   ]
(Where 0 is 'not at all satisfied' and 10 is 'completely satisfied')

Overall, how happy did you feel yesterday? [   ]
(Where 0 is 'not at all happy' and 10 is 'completely happy')

18 What is your height? _____ft _____ins or _____m _____cms

19 What is your weight? _____st _____lbs or _____kg

Marketing and Communications
By signing up to the 'Leeds Let's Get Active' programme, you are agreeing to receive marketing and communications to help you in taking steps for a healthier life. Once you have received your confirmation email, you will be able to opt out of future communications at any point.
By signing this application form, I agree to the terms and conditions of Leeds Let's Get Active project.

Signed: _____ Date: _____

Thank you for your time, Enjoy getting active!

Appendix E

Data processing agreement

E.1   Data processing agreement

LEEDS CITY COUNCIL

DATA PROCESSING AGREEMENT

THIS AGREEMENT is made

BETWEEN

The Parties

- Leeds City Council, (herein after called the "Data Controller") of Civic Hall, Calverly Street, Leeds, LS1 1UR
- *The University of Leeds* (herein after called the "Data Processor") of Leeds, LS2 9JT

Purpose

The purpose of the disclosure is to develop research and understanding on the impact, cost effectiveness and return on investment of the Leeds Let's Get Active (LLGA), a physical activity offer made by the Sport and Active Lifestyles service.

This will be done by allowing the Data Processor access to information with regards to "Leeds Let's Get Active" participants. This will be done to support in answering a number of specified research questions to produce anonymised reports and journal papers shared and published both locally and nationally to build the evidence base and local understanding.

Definitions

The following words and phrases used in this Agreement shall have the following meanings except where the context otherwise requires:

The expressions "Data", "Data Controller", "Data Processor", "Personal Data", "Sensitive Personal Data", "Processing", "Information Commissioner", "Data Subject Access" have the same meaning as in Sections 1, 2, and 6 of The Data Protection Act 1998, as amended by The Freedom of Information Act 2000.

"**Confidential Information**" means any information relating to the Data Controller's customers and prospective customers, current or projected financial or trading situations, business plans, business strategies, developments and all other information relating to the Data Controller's business affairs including any trade secrets, know-how and any information of a confidential nature imparted by the Data Controller to the Data Processor during the term of this Agreement or coming into existence as a result of the Data Processor's obligations, whether existing in hard copy form or otherwise, and whether disclosed orally or in writing. This definition shall include all Personal Data.

"**Services**" means the services to be provided by the Data Processor during the term of this Agreement.

1

Information provision

- *The information for processing within this Agreement relates to Leeds City Council's Sport and Active Lifestyles Service's data on its "Leeds Let's Get Active" participants. Individual participant identification should be anonymised. This data relates to:*
- *leisure centres used and/or community activities accessed;*
- *age of customer;*
- *leisure centre membership type;*
- *report code of activity undertaken;*
- *gender;*
- *postcode district;*
- *ethnicity;*
- *current employment status*
- *relationship status*
- *self-reported lifestyle risk factors including smoking status, alcohol consumption, eating habits, physical activity and wellbeing*
- *number of times they visited an activity.*
- *Long term condition(s) (self-reported)*
- *Data Processor may use this data for the time it takes to complete the purpose unless otherwise instructed by the Data Controller.*
- *The Data Processor will destroy the data (to the extent possible) on instruction from the Data Controller and not use it for any other purpose than this project, except that one copy may be kept by the Data Processor to the extent necessary to enable the student to complete their PhD (which may be placed on restricted access in the library of the University if requested) and for the purpose of defending itself against any claims arising in connection with this Agreement. It can only be used for work with Leeds City Council's Sport and Active Lifestyles Service as outlined in the purpose.*
- *Ownership of the Data shall at all times remain with the Data Controller*

Use, Disclosure and Publication

- The data will be used solely for the purpose outlined in this Agreement of supporting Leeds City Council's Sport and Active Lifestyles Service with its "Leeds Let's Get Active" project.
- Data will not be disclosed to any third party.
  The Data Processor will implement appropriate security measures to ensure no breach of the Data occurs whilst it holds the data
- *The Data Processor will not contact any Data Subject identified in the Data without the prior permission of the Data Controller*

Data Protection and Human Rights

The use and disclosure of any Personal Data shall be in accordance with the obligations imposed upon the Parties to this Agreement by the Data Protection Act 1998 and the Human Rights Act 1998. All relevant codes of

practice or data protection operating rules adopted by the Parties will also reflect the data protection practices of each of the parties to this Agreement. The Parties agree and declare that the information accessed pursuant to this Agreement will be used and processed with regard to the rights and freedoms enshrined within the European Convention on Human Rights. Further, the Parties agree and declare that the provision of information is proportional, having regard to the purposes of the Agreement and the steps taken in respect of maintaining a high degree of security and confidentiality.

The receipt by the Data Processor from any Data Subject of a request to access the Data covered by this Agreement must be reported immediately to the officer nominated representing the Data Controller, who will arrange the relevant response to that request.

Where the Data Processor receives a Notice under Section 10 of the Data Protection Act 1998 the Data Processor must report this immediately to the nominated officer representing the Data Controller who will arrange the relevant response.

If the Data Processor receives a request for information under the provisions of the Freedom of Information Act 2000 this must be sent immediately to the officer nominated within this Agreement as representing the Data Controller who will arrange the relevant response.

Where the Data Processor receives a Notice under Section 10 of the Data Protection Act 1998, the Data Processor will contact the officer nominated below to ascertain whether or not to comply with that Notice.

The following personnel are authorised by the Parties to assume responsibility for data protection compliance, notification, security, confidentiality, audit and co-ordination of subject rights and Freedom of Information:

The Data Processor shall give reasonable assistance as is necessary to the Data Controller in order to enable him to:

- Comply with request for subject access from the Data Subjects;
- Respond to Information Notices served upon him by the Information Commissioner;
- Respond to complaints from Data Subjects;
- Investigate any breach or alleged breach of the Act.

in accordance with his statutory obligations under the Data Protection Act 1998.

Nominated Post holder	Relevant LA
Rachel Brighton	Leeds City Council

On reasonable notice, periodic checks may be conducted by the Data Controller to confirm compliance with this Agreement.

3

Confidentiality

The Data Processor shall not disclose personal data to any person, other than those whose province it is to know the same for the Purpose, other than to the extent required under a court order, or pursuant to any order or requirement of a governmental or statutory body, provided that the Parties shall give notice in writing to the Data Controller of any disclosure they are required to make immediately they are aware of such a requirement.

The Data Processor shall ensure that any Individuals involved in the Purpose and to whom Data is disclosed under this Agreement are aware of their responsibilities in connection with the use of that Data and have confirmed so in writing.

For the avoidance of doubt, the obligations or the confidentiality imposed on the Parties by this Agreement shall continue in full force and effect after the expiry or termination of this Agreement.

Security

The Data Processor recognises that the Data Controller has obligations relating to the security of Data in his control under the Data Protection Act 1998. The Data Processor will continue to apply those relevant obligations as detailed below on behalf of the Data Controller during the term of this Agreement.

The Data Processor agrees to apply appropriate security measures, commensurate with the requirements of 7th principle of the Data Protection Act 1998 to the Data, which states that: "appropriate technical and organisation measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data". In particular, the Data Processor shall ensure that measures are in place to do everything reasonable to:
- make accidental compromise or damage unlikely during storage, handling, use, processing transmission or transport
- deter deliberate compromise or opportunist attack, and;
- promote discretion in order to avoid unauthorised access

The Data Processor will ensure that the personal data accessed is not used other than as identified within this Agreement, and that the Agreement is complied with.

The Data Controller reserves the right to undertake a review of security provided by any Data Processor and may request reasonable access during normal working hours to the Data Processor premises for this purpose. Failure to provide sufficient guarantees in respect of adequate security measures will result in the termination of this Agreement

4

The Data Processor undertakes not to use the services of any sub-contractors in connection with the processing of the Data without the prior written approval of the Data Controller.

Indemnity

In consideration of the provision of the Data for the Purpose the Data Processor undertakes to indemnify and keep indemnified the Data Controller against any direct liability, which may be incurred by the Data Controller as a result of the Data Processor's breach of this Agreement.

Provided that this Indemnity shall not apply:

(a) where the liability arises from information supplied by the Data Controller which is shown to have been incomplete or incorrect, unless the Data Controller establishes that the error did not result from any wilful wrongdoing or negligence on his part

(b) unless the Data Controller notifies the Data Processor as soon as possible of any action, claim or demand to which this indemnity applies, commits the Data Processor to deal with the action, claim or demand by settlement or otherwise and renders the Data Processor all reasonable assistance in so dealing;

(c) to the extent that the Data Controller makes any admission which may be prejudicial to the defence of the action, claim or demand.

In no event shall either Party be liable to the other for any consequential, incidental or indirect losses such as but not limited to loss of profit, revenue, contracts or the like.

Nothing in the Agreement limits or excludes the liability of either Party for death or personal injury caused by its negligence.

Disputes

In the event of any dispute or difference arising between the Parties out of this Agreement the Parties shall meet in an effort to resolve the dispute or difference in good faith.

The Parties will, with the help of the Centre for Effective Dispute Resolution, seek to resolve disputes between them by alternative dispute resolution. If the Parties fail to agree within 56 days of the initiation of the alternative dispute resolution procedure, then the Parties shall be at liberty to commence litigation.

Term, Termination and Variation

The Data Controller may at any time by notice in writing terminate this Agreement forthwith if the Data Processor is in material breach of any obligation under this Agreement.

5

Either Party may terminate this Agreement by giving 30 days notice in writing to the other Party.

The Data Controller will have the final decision on any proposed variation to this Agreement. No variation of the Agreement shall be effective unless it is contained in a written instrument signed by both Parties and annexed to this Agreement.

Miscellaneous

This Agreement acts in fulfilment of part of the responsibilities of the Data Controller as required by paragraphs 11 and 12 of Schedule 1, Part II of the Data Protection Act 1998.

If any provision of this Agreement is held by a Court of competent jurisdiction to be invalid or unenforceable, such invalidity or unenforceability shall not affect the remaining provisions of this Agreement, which shall remain in full force and effect.

The validity, construction and interpretation of the Agreement and any determination of the performance which it requires shall be governed by the Laws of England and the Parties hereby submit to the exclusive jurisdiction of the English Courts.

Signed on behalf of Leeds City Council
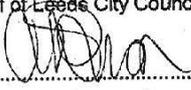
.... J M ALLMAN ....
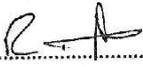
In the presence of .... RACHEL BRIGHTON ....

Signed on behalf of The University of Leeds
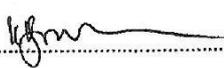
KATHY BROWNRIDGE
.... DIRECTOR ....
RESEARCH & INNOVATION SERVICE

In the presence of ..Karen Ogier, Contracts Manager...............

Signed as seen and agreed to by the individuals receiving the data;
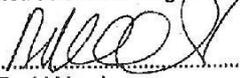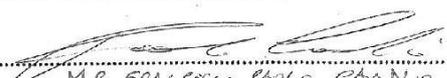
....
Mr David Meads   MR FRANCESCO PAOLO CAPALDO
                              (PHD STUDENT)

NB – To be signed by the student once appointed.

6

Appendix F

Probability of service use

F.1   Probability of using the service at least once

Characteristi	Category	Unadjusted RR	P-value	Adjusted RR	P-value
Age	16-40	1.00	NA	1.00	NA
	41-64	0.992 (1.078-1.084)	0.344	0.937 (0.872-	0.079
	Over 64	1.088 (1.060-1.118)	<0.001	1.085 (0.947-	0.241
Gender	Female	1.00	NA	1.00	NA
	Male	1.085 (1.068-1.101)	<0.001	1.099 (1.029-	0.005
IMD	Non-deprived LSOA	1.00	NA	1.00	NA
	Top 20% score	0.882 (0.854-0.910)	<0.001	1.013 (0.929-	0.765
	Top 10% score	1.060 (1.023-1.098)	0.001	NA	NA
	Top 3% score	0.964 (0.882-1.053)	0.421	NA	NA
PA category	Inactive	1.00	NA	1.00	NA
	Insufficiently active	1.119 (1.091-1.147)	<0.001	1.121(1.035-	0.005
	Moderately active	1.199 (1.067-1.232)	<0.001	1.181 (1.079-	<0.001
	Active	1.085 (1.148-1.222)	<0.001	1.074 (0.955-	0.230

Notes: see table 2 + NA=not applicable, PA= physical activity, RR=risk ratio

F.2   Probability of higher level of service use

Characteristic	Category	4° quartile		3° quartile		2° quartile	
		Unadjusted RRR	Adjusted RRR	Unadjusted RRR	Adjusted RRR	Unadjusted RRR	Adjusted RRR
Age	16-40	1.00	1.00	1.00	1.00	1.00	1.00
	41-64	**1.929*** (1.810-2.057)	**2.460*** (2.810-3.343)	**1.213*** (1.139-1.292)	1.197 (0.940-1.525)	**1.091** (1.022-1.165)	1.151 (0.911-1.455)
	Over 64	**3.472*** (3.132-3.848)	**2.233** (1.319-3.780)	**1.427 *** (1.275-1.597)	1.358 (0.851-2.170)	**1.212** (1.076-1.364)	0.948 (0.570-1.576)
Gender	Female	1.00	1.00	1.00	1.00	1.00	1.00
	Male	**1.370*** (1.293-1.452)	**2.140*** (1.610-2.844)	1.009 (0.952-1.069)	1.294* (1.036-1.616)	0.956 (0.900-1.014)	1.089 (0.875-1.356)
IMD	Non-dep. LSOA	1.00	1.00	1.00	1.00	1.00	1.00
	Top 20% score	**0.812*** (0.725-0.911)	1.191 (0.818-1.733)	**0.892*** (0.780-0.995)	1.008 (0.750-1.355)	0.976 (0.875-1.089)	1.081 (0.814-1.436)
	Top 10% score [a]	**1.745*** (1.522-1.999)	**1.821*** (1.584-2.094)	**1.430** (1.244-1.645)	**1.388*** (1.205-1.597)	**1.261** (1.089-1.461)	**1.227** (1.058-1.422)
	Top 3% score [a]	**1.876*** (1.337-2.635)	**2.188*** (1.549-3.090)	**1.891** (1.354-2.640)	**1.899*** (1.3-2.660)	1.193 (0.818-1.741)	1.191 (0.814-1.741)
Physical activity category	Inactive	1.00	1.00	1.00	1.00	1.00	1.00
	Ins. active	1.052 (0.960-1.152)	1.298 (0.904-1.864)	1.016 (0.929-1.111)	0.994 (0.764-1.294)	1.009 (0.922-1.106)	0.912 (0.707-1.175)
	Mod. active	1.121* (1.013-1.241)	1.114 (0.744-1.669)	1.011 (0.915-1.117)	0.872 (0.646-1.178)	0.904 (0.815-1.002)	0.865 (0.647-1.156)
	Active	**1.300*** (1.157-1.463)	1.224 (0.752-1.995)	0.981 (0.870-1.105)	0.780 (0.529-1.151)	0.943 (0.834-1.066)	0.708 (0.482-1.041)
PRE	Previous	1.00	1.00	1.00	1.00	1.00	1.00
	New member	**0.440*** (0.408-0.476)	0.506(0.255-1.007)	**0.767***(0.706-0833)	0.676 (0.364-1.257)	0.892* (0.817-0.974)	0.800 (0.420-1.522)

Notes: see table 2 + [a] adjusted for age, gender, PAcat0 and PRE (data available) *0.01 -0.05 Boldface $>$0.001 - 0.01; Boldface* <0.001, IMD=Index of Multiple Deprivation, PRE=gym member before LLGA started

Appendix G

Physical activity transition probabilities: last observation carried forward

Transition matrices – survey measure

NON - DEPRIVED				
	Inactive	Insufficient	Mod Act	Active
Inactive	99.268%	0.732%	0.000%	0.000%
Insufficient	0.460%	98.691%	0.849%	0.000%
Mod Act	0.000%	0.754%	98.643%	0.603%
Active	0.000%	0.000%	0.482%	99.518%

DEPRIVED				
	Inactive	Insufficient	Mod Act	Active
Inactive	99.268%	0.732%	0.000%	0.000%
Insufficient	0.452%	98.684%	0.864%	0.000%
Mod Act	0.000%	0.656%	98.651%	0.693%
Active	0.000%	0.000%	0.639%	99.361%

Transition matrices – card swipe measure

NON - DEPRIVED				
	Inactive	Insufficient	Mod Act	Active
Inactive	99.415%	0.585%	0.000%	0.000%
Insufficient	0.090%	99.440%	0.470%	0.000%
Mod Act	0.000%	0.075%	99.451%	0.475%
Active	0.000%	0.000%	0.000%	100.000%

DEPRIVED				
	Inactive	Insufficient	Mod Act	Active
Inactive	99.415%	0.585%	0.000%	0.000%
Insufficient	0.085%	99.415%	0.501%	0.000%
Mod Act	0.000%	0.073%	99.440%	0.487%
Active	0.000%	0.000%	0.000%	100.000%

Appendix H

Probability of second-stage survey response

Multivariate probit coefficients for NAD response

Variable	Category	Coefficient	p-value
NAD	0	Reference category	
	1	**0.148**	0.003
	2	0.008	0.875
	3	0.077	0.172
	4	0.082	0.267
	5	0.005	0.948
	6	-0.043	0.774
	7	-0.009	0.928
Gender	Female	Reference category	
	Male	-0.037	0.312
Age	16-40	Reference category	
	41-64	**0.328**	<0.001
	>64	**0.319**	<0.001
IMD	Non-deprived LSOA	Reference category	
	Top 20%	-0.044	0.408
	Top 10%	0.080	0.395
Cohort	1	Reference category	
	2	**0.767**	<0.001
PRE	Previous member	Reference category	
	New member	**0.222**	0.011
LLGA	No LLGA attendance	Reference category	
	At least 1 LLGA session	**0.553**	<0.001
OUT	No attendance outside LLGA	Reference category	
	At least 1 session outside	**0.140**	<0.001

Note: NAD=number of active days, IMD=Index of Multiple Deprivation

Appendix I

City Council financial audit reports

LLGA Year 1 spend - 010413 - 310314

Activity	Total Spend	Comments
Staffing		
Project lead - seconded staff	38878	value of staff time before appointment to LLGA lead post
Project lead	22143.95	started mid- way through financial year
Staff badges	9.9	
Staff telephones	344.69	
Staff travel	239.19	
In-kind staffing contribution from service	225760	This is the value of the staffing support provided to the project across the wholeservice
Loss of income	125545	includes gym, swim and gym inductions
Community programme		
Coaching staff	2872.21	
Equipment	228.22	
venue hire	300	
promo	70	
training	440	
Marketing and communication		
Promotional materials	12171.73	
Partner engagement events	737.77	
Microsite	17000	design and management of
Launch campaign - radio	2273	
Launch campaign - bus shelters	10988	
Launch campaign - other	11230.4	
Research	24650	
Administration		
Postal fees	6.22	

LLGA Year 2 spend - 010414 - 310315

Activity	Total Spend	Comments
Staffing		
Project lead and project coordinator posts	65632.42	
Staff travel	558.47	
In-find staff contribution from service	160000	the contribution made by staff across the whole service to the project delivery
Loss of income		
Leeds City Council	249608	for gym, swim and inductions
Bramley Baths	15600	external partner who we pay a loss of income fee to for doing LLGA
Community programme		
Coaching staff	11411.59	
Equipment	1634.25	
venue hire	1203	
promo	1165.4	
training	324.3	
Marketing and communication		
Promotional materials	11417.19	leaflets, art design, posters etc
Partner engagement events	393.87	
Promotional film	6537	
Social media promotions	93.86	
Research	24650	
Administration		
IT - fee to cascade for reporting	2970	

LLGA Year 3 spend - 010414 - 310315

Activity	Total Spend	Comments
Staffing		
Project lead and project coordinator posts	29075.1	both staff members left in this financial year
Staff travel	623.46	
In-find staff contribution from service	80000	contribution from all staff across the service to delivery of LLGA
Staff costs to cover vacant LLGA lead	13365.1	
staff phones	49.82	
Loss of income		
Leeds City Council	249590	
Bramley Baths	10400	
Community programme		
Coaching staff	4972.76	
Community delivery - non LCC coaches	9646.73	
Equipment	391.45	
venue hire	1330	
promo	88.59	
Music licences	75	
Marketing and communication		
Promotional materials	8244.34	
Partner engagement events	75	
Social media promotions	7.03	
Research		
Leeds Beckets University	24650	
SMG insite	10000	
Focus groups	140	
Total	442724.38	

LLGA Year 4 spend - 010416 - 311216

Activity	Total Spend	Comments
Staffing		
Project lead and project coordinator posts	12,925.06	Project lead recruited Sept 2016
Staff travel	41.76	
In-Kind staff contribution from service	106,000	contribution from all staff across the service to delivery of LLGA
Staff phones	194.54	
Loss of income		
Leeds City Council	191200	
Bramley Baths	7800	
Community programme		
Coaching staff	4776.69	
Community delivery - non LCC coaches	8572.5	
Equipment	100.99	
Venue hire	620	
Marketing and communication		
Promotional materials	4,630.53	
Research		
Leeds Beckets University	16,433	
Total	353,295.07	

Appendix J

Decision-analytic model parameters and settings

J.1   Decision-analytic model parameters and settings

Model parameters	Health state	Parameter	Source / Method	Distribution
Annual prob ACT	T2D	0.002 (0.002)	Joseph et al. (2010)	Beta
	CHD	0.008 (0.0005)	Frew et al. (2014)	Beta
	STR	0.011 (0.0031)	Frew et al. (2014)	Beta
	CRC	0.003 (0.003)	Frew et al. (2014)	Beta
	BRC	0.011 (0.001)	Frew et al. (2014)	Beta
	DEP	0.011 (0.0106)	National Institute for Health and Care Excellence (2015)	Beta
	FRA	0.023 (0.023)	Fried et al. (2001)	Beta
RR INA	T2D	1.700 (1.7)	Roux et al. (2008)	LogNormal
	CHD	1.500 (1.5)	Roux et al. (2008)	LogNormal
	STR	1.300 (1.3)	Roux et al. (2008)	LogNormal
	CRC	1.600 (1.6)	Roux et al. (2008)	LogNormal
	BRC	1.300 (1.3)	Roux et al. (2008)	LogNormal
	DEP	1.150 (1.15)	Meng and D'Arcy (2013)	LogNormal
	FRA	1.429 (1.43)	McPhee et al. (2016)	LogNormal
RR INS	T2D	1.525 (1.52)	Linear interpolation	LogNormal
	CHD	1.375 (1.137)	Linear interpolation	LogNormal
	STR	1.225 (1.225)	Linear interpolation	LogNormal
	CRC	1.450 (1.45)	Linear interpolation	LogNormal
	BRC	1.225 (1.225)	Linear interpolation	LogNormal
	DEP	1.113 (1.11)	Linear interpolation	LogNormal
	FRA	1.321 (1.32)	Linear interpolation	LogNormal
RR MOD	T2D	1.292 (1.29)	Linear interpolation	LogNormal
	CHD	1.208 (1.208)	Linear interpolation	LogNormal
	STR	1.125 (1.125)	Linear interpolation	LogNormal
	CRC	1.250 (1.25)	Linear interpolation	LogNormal
	BRC	1.125 (1.125)	Linear interpolation	LogNormal
	DEP	1.063 (1.063)	Linear interpolation	LogNormal
	FRA	1.179 (1.179)	Linear interpolation	LogNormal
RR IMD	T2D	1.250 (0.041)	Sharma et al. (2016)	LogNormal
	CHD	1.294 (1.29)	Bajekal et al. (2012)	LogNormal
	STR	1.400 (1.4)	Bray et al. (2018)	LogNormal
	CRC	1.100 (1.1)	Cancer Research UK (2011)	LogNormal
	BRC	0.860 (0.86)	Cancer Research UK (2011)	LogNormal
	DEP	1.170 (0.296)	Walters et al. (2012)	LogNormal
	FRA	1.100 (0.11)	Curtis et al. (2016)	LogNormal

RR death	T2D	1.850 (0.332)	Nwaneri et al. (2012)	LogNormal
	CHD	1.900 (0.161)	Vlachopoulos et al. (2012)	LogNormal
	STR	1.900 (0.161)	Vlachopoulos et al. (2012)	LogNormal
	CRC	1.449 (1.45)	Cancer Australia (2016)	LogNormal
	BRC	1.320 (0.041)	Christiansen et al. (2011)	LogNormal
	DEP	1.520 (0.036)	Cuijpers et al. (2014)	LogNormal
	FRA	2.700 (0.74)	Kulmala et al. (2014)	LogNormal
Utility decrements	T2D	0.062 (0.06)	Sullivan and Ghushchyan (2016)	Gamma
	CHD	0.056 (0.06)	Gulliford et al. (2014)	Gamma
	STR	0.101 (0.101)	Gulliford et al. (2014)	Gamma
	CRC	0.038 (0.038)	Gulliford et al. (2014)	Gamma
	BRC	0.015 (0.015)	Sullivan et al. (2005)	Gamma
	DEP	0.130 (0.13)	Gulliford et al. (2014)	Gamma
	FRA	0.177 (0.18)	Lin et al. (2011)	Gamma
Utility values IMD NON-DEPRIVED	INA	0.935 (0.0221)	HSE 2014 data analysis	Beta
	INS	0.985 (0.0218)	HSE 2014 data analysis	Beta
	MOD	0.997 (0.0223)	HSE 2014 data analysis	Beta
	ACT	0.982 (0.0219)	HSE 2014 data analysis	Beta
Utility values IMD DEPRIVED	INA	0.935 (0.0221)	HSE 2014 data analysis	Beta
	INS	0.979 (0.0228)	HSE 2014 data analysis	Beta
	MOD	0.981 (0.0239)	HSE 2014 data analysis	Beta
	ACT	0.986 (0.0225)	HSE 2014 data analysis	Beta
Treatment and management costs	T2D	£ 1,363	Frew et al. (2014)	Gamma
	CHD1	£ 3,489	Frew et al. (2014)	Gamma
	CHD2	£ 105	Frew et al. (2014)	Gamma
	STR1	£ 9,630	Frew et al. (2014)	Gamma
	STR2	£ 2,396	Frew et al. (2014)	Gamma
	CRC	£ 9,999	Frew et al. (2014)	Gamma
	BRC	£ 9,091	Frew et al. (2014)	Gamma
	DEP	£ 139	Thomas and Morris (2003)	Gamma
	FRA	£ 3,351	McNamee et al. (1999)	Gamma

Notes: HSE=Health Survey for England; IMD=Index of Multiple Deprivation status; INA=inactive; INS=insufficiently active; MOD=moderately active; ACT=active; T2D=Type II Diabetes; CHD1=Coronary Heart Disease, first year from event; CHD2=Coronary Heart Disease, second and subsequent years; STR1=Stroke, first year from event; STR2=Stroke, second and subsequent years; CRC=Colorectal Cancer; BRC=Breast Cancer; DEP=Depression; FRA=Frailty syndrome, RR=Relative Risk.

Appendix K

Leeds City Council: Executive Board report

Leeds
CITY COUNCIL

| Report author: Mark Allman |
| Tel: 2478323 |

LeeReport of: Director of City Development and Director of Public Health

Report to: Executive Board

Date: 24ᵗʰ April 2013

Subject: Leeds Lets Get Active

Are specific electoral Wards affected? If relevant, name(s) of Ward(s):	☐ Yes	☒ No
Are there implications for equality and diversity and cohesion and integration?	☒ Yes	☐ No
Is the decision eligible for Call-In?	☒ Yes	☐ No
Does the report contain confidential or exempt information? If relevant, Access to Information Procedure Rule number: Appendix number:	☐ Yes	☒ No

Summary of main issues

1.   Executive Board were previously informed of work in Birmingham in providing free activities for residents at selected times and venues within the city, funded via public health and with great success.

2.   Shortly after the Executive Board in September, Sport England announced a new £5m national health pilot fund "Get Healthy, Get into Sport". The fund was designed to support projects that can demonstrate health gains through sport and physical activity and, vitally, provide a robust evidence base. Leeds was one of only 16 projects (from over 280 applicants) that were asked to develop a detailed bid. A formal bid was submitted on February 8ᵗʰ 2013 and we received confirmation that the project had been successful in securing this funding on the 19ᵗʰ March 2013. The Sport England funding of £500k is being matched in cash terms by Leeds City Council

(Public Health) together with considerable "in kind" support. The Leeds scheme will be known as "Leeds Lets Get Active" and will focus on providing a universal free offer.

Recommendations

Executive Board is recommended to:

(i)   Note the contents of the report and support the project.

(ii)   Grant approval to the Director of City Development to accept the Sport England grant funding award of £500,000.

(iii)   Request a report at the end of the project evaluating the outcomes.

1   Purpose of this report

1.1   To provide further information relating to the City Councils Leeds Lets Get Active bid to Sport England's "Get into healthy, Get into Sport" health pilot programme.

1.2   To seek retrospective support for the Leeds Lets Get Active bid and seek approval to accept a grant offer.

2   Background information

2.1   The Head of Sport and Active Lifestyles has been closely engaged with Sport England nationally in the development of their current funding strategy. One of the funding strands that Sport England and Local Authorities were keen to explore further was around the contribution sport and "being active" makes to public health outcomes.

2.2   Services that increase physical activity have the potential to reduce all-cause mortality and improve life expectancy. Even relatively small increases in physical activity are associated with some protection against chronic diseases, improved mental health and an improved quality of life. Physical activity can also save money by significantly easing the burden of chronic disease on the health and social care services and has the potential to reduce transport costs through the promotion of active travel. *CMO's 'Start Active, Stay Active'*. For example, a brisk walk every day in your local park can reduce the risk of heart attacks by 50%, strokes by 50%, diabetes by 50%, fracture of the femur by 30%, colon cancer by 30% and alzheimers by 25% (Dr William Bird 2002).

2.3   Sport England launched its 'Get Healthy, Get into Sport' funding stream in September 2012. Leeds City Council and NHS Leeds/Public Health submitted a joint proposal based on an adaptation of the Birmingham Be Active model. The proposal is divided into two key strands. Firstly a core offer based on evaluating the impact of targeted free use of leisure centres (Bodyline gyms and swimming between 1 and 2 hours every day), focussing in areas of greatest health inequality. Secondly this work was to be supported by further interventions in community settings and improved health referral routes via the health sector and other customer contact points.

2.4   On the 19th March 2013 it was confirmed to Leeds that the bid submission has successfully secured funding through the Sport England 'Get Healthy, Get into Sport fund. Sport England will be funding the project to a value of £500,000 and this is being matched in cash terms by Leeds City Council (Public Health) together with considerable "in kind" support. Leeds will be working closely with an academic partner to evaluate the project which will run from October 2013 to March 2015. Progress and impact will be reported via the appropriate channels within public health with the ambition of mainstreaming the funding should the outcomes be met.   The Leeds scheme will be known as Leeds Lets Get Active (LLGA). This links it to the 'Leeds Lets Change' campaign.

3   Main Issues

3.1   LLGA seeks to explore methods to remove barriers that exist for the least active people in Leeds in relation to participating in sport and physical activity. It hopes to initiate a change in culture whereby inactive people take small steps to being active, feeling encouraged to take part in sport and physical activity in an environment where they feel welcome and comfortable. The ultimate aim is to help reduce the significant health inequalities that exist in the city.   Furthermore by getting people doing some activity it is anticipated (through the right interventions) that they can progress into a range of sports (hence Sport England's interest). The project will test the barriers to participation (getting the inactive active) and what methods most effect behaviour change. The bid is based on 3 key strands, namely 1) a core sport / fitness activity offer in leisure centres; 2) a community multi-sport offer and 3) a behaviour change intervention within the Bodyline Access Scheme. More detail is provided below on each of these areas:

3.1.1   **Strand 1: Testing the impact of free/discounted use of Leeds City Council leisure centres for selected sport and fitness activities**, at selected times, daily, for all Leeds residents (universally targeted).

- The offer will be greatest in areas of the city where activity levels are lowest and health inequalities are highest

- The offer in leisure centres will typically be one free hour every day (off peak) with an additional hour per day for 4 leisure centres that serve the most deprived areas of the city, namely, John Charles Centre for Sport, Armley, Fearnville and Middleton Leisure centres.

- Activities to include gym and swim, except at Middleton Leisure centre where a specific programme will be developed

3.1.2   **Strand 2: Testing the impact of free / discounted use of community multi-sport sessions**

- The offer will be greatest in areas of the city where activity levels are lowest and health inequalities are highest

- Activities to include Running, Walking for Health and family multi-sport activities

- The programmes will be delivered in blocks of 10 – 12 weeks. In total there will be 102 blocks of activity over the life time of the project. The delivery will mainly take place in parks.

3.1.3 **Strand 3: Testing the impact of behaviour change interventions on the uptake of the Bodyline Access Scheme**

- Extending the existing Bodyline Access Scheme (based on £5 for 3 months worth of activity that includes, swimming/Bodyline gyms/classes at off peak times including weekends), linking to NHS health check via GP's and healthy lifestyle services.

- Developing a more integrated process for health professionals into LLGA that supports people 'who could benefits from doing more activity'.

- An evidence based package of support for the new user that will aid their behaviour change

3.2 The projects' success will be judged by a range of measures including for example, helping Leeds to meet its ultimate ambition of being "the most active big city", as well as reducing health inequalities, demonstrating the value of Sport and Active Lifestyles in supporting health outcomes (all age all cause mortality, cardiac conditions, weight loss, functional health, cancer, diabetes) and creating a strong enough case for future funding support. If successful it is envisaged that the project will grow and potentially the free offer will be expanded both in quantity and in relation to the range of activities on offer.

3.3 Following on from this the main aims of the project are summarised below:

- To increase the activity levels of those who are inactive in the city, especially in areas that have the highest health inequalities in adults and young people.
- To understand the barriers to being active for adults and young people
- To better understand what methods can be successfully deployed to move people from being inactive to undertaking 30 minutes of activity per week
- Establish better links with health partners including commissioners and healthcare partners

3.4 Attached as Appendix 1 is the research framework for the project. A research partner will be contracted to work alongside the council to support with the delivery of the research methodology. This partnership will explore the value of using various research and evaluation techniques of both a qualitative and quantitative nature and will build on studies already undertaken e.g. Birmingham Be Active (BCC and Matrix) /Fit for the Future (DOH 2009 -2010). The research methodology will influence project development and, therefore, the research partner will form part of the detailed project team.

3.5 LLGA will make free and discounted sessions conditional on carrying a Leeds Active card. This is essential as it will allow data to be compiled about those customers who are new and those who are already engaged. Sport England's main aim is to provide a strong evidence base of impact. New participants on disability or income related

benefits will also be promoted to and offered the additional feature of the Leeds 'Extra' card to encourage activity beyond what is freely available. In addition all new participants will be asked to complete a questionnaire at the beginning of the programme that will assess physical activity levels prior to the start of the scheme.

3.6   It is proposed that the initial targeted marketing campaign will promote LLGA with a call to action to apply for your new Leeds Active Card, providing access to free health and fitness opportunities at your local leisure centre and in your local community. A combination of traditional and digital techniques are to be applied, ensuring that the chosen techniques are relevant and appropriate to the intended target market. A key aim of this programme is to address inequalities in sports participation, and we will be able to use profiling to identify people who are more likely to be physically inactive and more at risk of developing medical conditions in future. A targeted approach to the marketing and communication will be vital to the success of the scheme as it will ensure the promotional campaign is directly focussed at the people the scheme aims to benefit – those who are inactive.

3.7   The project will be managed through a joint partnership with health, sport and active recreation professionals. A Project Board will be established and report both to Sport Leeds partnership and to the Health and Wellbeing Board and / or associated health boards. Funding is available to support staffing, this includes a full time project lead to oversee the scheme and a part time (.5) coordinator to manage the Bodyline Access programme and to oversee all the participant support programmes (i.e. 1-2-1 goal setting, champions scheme).

3.8   The funding award from Sport England is dependant on the following conditions:

There are 5 standard conditions that all successful Get Healthy funded projects will need to meet:

- funding will be awarded for Year 1 and then Year's 2 and 3 will be awarded in principle linked to tangible outcomes/outputs for each project
- Sport England will not release the first payment until we have written confirmation of all partnership funding
- the project will not involve any sport that is not recognised by Sport England
- no element of the award will be used to cover the redundancy costs of any at risk posts linked to the delivery of your project
- an evaluation plan must be submitted for Sport England's approval

In addition the following bespoke project conditions are attached to the Leeds Lets Get Active Project:

- Alongside Sport England's standard monitoring information included in the award offer they would look for evidence in January 2014 that their contribution will be focused on attracting new users (accepting that a proportion of this will support existing and those diverted from other sessions), that there is some evidence of Leeds success in attracting new users as well as figures on participation where available
- As part of the discussion around Year 2 Sport England will also ask for a sustainability plan with an operational budget for the following years.

- Sport England colleagues (including Facilities & Planning) will be an integral part of the project management

3.9 Now that funding has been confirmed for the project through Sport England and public health the next steps include recruitment of the project lead, engagement of a research partner, development of the free offer product in leisure centres and community and initiation of the engagement plan. The first LLGA project board is due to take place on the 23rd April 2013.

3.10 The LLGA project presents an opportunity to reinforce the value of being active with our young people, something of keen interest to the Youth Mayor. Furthermore there are opportunities for LLGA to feature as part of a more coherent sport and physical activity offer for young people, for example by connecting up work associated with the youth review as well as recent Government announcements outlining significant investment in primary school PE and Sport. The project will also be developed alongside other key initiatives that focus on young people and adults as part of the wider Olympic legacy programme including for example, major events (e.g. Rugby league world cup/ Tour de France) , National Governing Body "Place Pilot", sport legacy fund and community access to school sport facilities.

4.   Corporate Considerations

4.1   Consultation and Engagement

4.1.1 Leeds Lets Get Active has been developed in partnership with Public Health and addresses priorities identified through the JSNA and the Sport England Active People survey. A public consultation took place from December 2012 to January 2013 to identify the key barriers to participation for inactive people and to collate views on how these could be overcome. SportLeeds (the city partnership for sport and active lifestyles) have been consulted on the development of the proposal on an ongoing basis. The sustainable economy and culture scrutiny board (9th April 2013) have also received details of the scheme as part of their wider enquiry into the role of Leisure and Culture in supporting the delivery of improved public health outcomes. There will be ongoing consultation as the project develops, including key stakeholders groups at both a city and local level as well as ward members in a effort to help reach the most inactive people.

4.2   Equality and Diversity / Cohesion and Integration

4.2.1 These proposals have been screened for issues on Equality, Diversity, Cohesion and Integration (EIA screening attached as an appendix). In general, such considerations are integral to this whole report as one of the major aims of the proposals is to narrow health inequality, a key council objective. As well as offers in the community, the proposed 18 month pilot offers free off-peak access to a swim or gym session for at least one hour every day in all leisure centres, two at those in areas of highest deprivation. Those currently unable to afford swimming and gyms should benefit most, wherever in Leeds they live. This may particularly benefit those on low incomes, minority ethnic groups and older people.

4.3   Council policies and City Priorities

4.3.1 The proposals aim to narrow health inequality, a major council objective, by encouraging more people to become more physically active, particularly those in areas of higher deprivation where activity levels and life expectancy are lower than the city's average.

4.3.2 The overarching vision for 2030 is that Leeds will be the best city in the UK. This means all Leeds' communities will be successful, including those who are currently less active and suffer poorer healthy life expectancy.

4.3.3 City Development has as a priority to "Develop the city's cultural events and facilities including changes to sports centres and libraries", and a key performance measure is "To maintain visits to sports centres". This report directly addresses these priorities.

4.4 Resources and value for money

4.4.1 Swimming pools and gyms carry significant costs to build, maintain and staff. Fee paying customers (casual, memberships, schools and clubs) currently cover a high proportion of the revenue cost of running leisure centres, so that the £6.2m managed budget in Sport is only 1.1% of the Council's total spend, and comparatively low compared to other comparable Local Authorities.

4.4.2 These proposals should be neutral to the council's budget in 2013/14 and 2014/15. New expenditure and income lost totalling £1,000,000 is being fully funded by Public Health and Sport England with £500,000 each. The 'in-kind' support worth £320,000 anticipated from officers in Sport Development and Facilities comprises work from existing employees who would otherwise be providing similar services.

4.5 Legal Implications, Access to Information and Call In

4.5.1 The provision of sport services by councils and their pricing or subsidy is not subject to statute so the main legal criteria is that these proposals are reasonable.

4.5.2 The decision is eligible for call-in.

4.6 Risk Management

4.6.1 The main financial risk is that the free offer diverts more paying customers than anticipated, widening the loss of income and reducing the space in pools for previously inactive newcomers. This would increase the cost and reduce the effect of the free swim part of the offer and it might have to be curtailed early to avoid loss to the council. To manage the risk the income loss and numbers of new participants will be monitored weekly for any disproportionate loss of income.

4.6.2 The main policy risk is that this pilot produces an expectation of free access to high cost facilities and activities at a public subsidy that cannot be sustained. To mitigate this risk, efforts will be made to offer additional paid sessions to new customers and to build up evidence of the benefits of the offer, so as to encourage future funding or sponsorship.

5. Conclusions

5.1   The LLGA projects provides an exciting opportunity to test the effectiveness of price discounting on participation and therefore health outcomes. The targeted nature of the project within a universal offer will provide a unique insight into behaviour change.

6.   Recommendations

Executive Board is recommended to:

(i)   Note the contents of the report and support the project.

(ii)   Grant approval to the Director of City Development to accept the Sport England grant funding award of £500,000.

(iii)   Request a report at the end of the project evaluating the outcomes.

Appendix L

Prospero review protocol

UNIVERSITY *of York*
Centre for Reviews and Dissemination

NHS
National Institute for
Health Research

PROSPERO International prospective register of systematic reviews

Review title and timescale

1   Review title
Give the working title of the review. This must be in English. Ideally it should state succinctly the interventions or exposures being reviewed and the associated health or social problem being addressed in the review.
Economic evaluations of physical activity promotion for primary prevention: A systematic literature review of methods and results

2   Original language title
For reviews in languages other than English, this field should be used to enter the title in the language of the review. This will be displayed together with the English language title.

3   Anticipated or actual start date
Give the date when the systematic review commenced, or is expected to commence.
01/09/2016

4   Anticipated completion date
Give the date by which the review is expected to be completed.
30/06/2017

5   Stage of review at time of this submission
Indicate the stage of progress of the review by ticking the relevant boxes. Reviews that have progressed beyond the point of completing data extraction at the time of initial registration are not eligible for inclusion in PROSPERO. This field should be updated when any amendments are made to a published record.

The review has not yet started   ✕

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	No
Data extraction	Yes	No
Risk of bias (quality) assessment	No	No
Data analysis	No	No

Provide any other relevant information about the stage of the review here.

Review team details

6   Named contact
The named contact acts as the guarantor for the accuracy of the information presented in the register record.
Francesco Paolo Candio

7   Named contact email
Enter the electronic mail address of the named contact.
umfpc@leeds.ac.uk

8   Named contact address
Enter the full postal address for the named contact.
LS4 2EN

9   Named contact phone number
Enter the telephone number for the named contact, including international dialing code.
0113 3430833

10   Organisational affiliation of the review
Full title of the organisational affiliations for this review, and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.

UNIVERSITY *of York*
Centre for Reviews and Dissemination

NHS
National Institute for
Health Research

University of Leeds

Website address:
http://www.leeds.ac.uk/

11   Review team members and their organisational affiliations
Give the title, first name and last name of all members of the team working directly on the review. Give the organisational affiliations of each member of the review team.

Title	First name	Last name	Affiliation
Mr	Francesco Paolo	Candio	University of Leeds

12   Funding sources/sponsors
Give details of the individuals, organizations, groups or other legal entities who take responsibility for initiating, managing, sponsoring and/or financing the review. Any unique identification numbers assigned to the review by the individuals or bodies listed should be included.
Independent research produced under the terms of the White Rose PhD Studentship Network scheme as part of the National Institute for Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC).

13   Conflicts of interest
List any conditions that could lead to actual or perceived undue influence on judgements concerning the main topic investigated in the review.
Are there any actual or potential conflicts of interest?
None known

14   Collaborators
Give the name, affiliation and role of any individuals or organisations who are working on the review but who are not listed as review team members.

Title	First name	Last name	Organisation details

Review methods

15   Review question(s)
State the question(s) to be addressed / review objectives. Please complete a separate box for each question.
1. What methodological approaches for economic evaluation have been used? 2. What major modelling assumptions have been made to determine effects on health? 3. Do physical activiy promotion programmes offer value for money? 4. If and how methodological challenges have been addressed?

16   Searches
Give details of the sources to be searched, and any restrictions (e.g. language or publication period). The full search strategy is not required, but may be supplied as a link or attachment.
Search strategy The following generic and specialised electronic databases will be searched for relevant studies in both the published and grey literature: • through OVID: o MEDLINE o EMBASE • Cochrane Library • through EBSCO: o SportDiscuss o EconLit A search strategy will consist of free-text and MeSH terms related to the four following concepts: (1) physical activity, (2) behaviour/lifestyle, (3) economic evaluation, (4) economic and decision model. Terms relating to intervention setting or type will not be included. This is justified as eligible papers may be missed in the search if intervention terms are included. Although this will undoubtedly produce a high yield of studies from the search, the sensitivity of the search (and therefore comprehensiveness of the systematic review) will be improved. As the use of technical terms for indexing international literature in databases is often inconsistent, I will define a search strategy for MEDLINE that will be then adapted as necessary to take into account of the differing indexing terms across the other databases. There will be no limits on the publication date. All records published in the English language will be included.

17   URL to search strategy
If you have one, give the link to your search strategy here. Alternatively you can e-mail this to PROSPERO and we will store and link to it.

I give permission for this file to be made publicly available

UNIVERSITY *of York*
Centre for Reviews and Dissemination

No

NHS
National Institute for
Health Research

UNIVERSITY *of York*
Centre for Reviews and Dissemination

NHS
National Institute for
Health Research

18   Condition or domain being studied
Give a short description of the disease, condition or healthcare domain being studied. This could include health and wellbeing outcomes.
Physical activity promotion

19   Participants/population
Give summary criteria for the participants or populations being studied by the review. The preferred format includes details of both inclusion and exclusion criteria.
Individuals from the general population

20   Intervention(s), exposure(s)
Give full and clear descriptions of the nature of the interventions or the exposures to be reviewed
Economic evaluations on physical activity promotion programmes / interventions

21   Comparator(s)/control
Where relevant, give details of the alternatives against which the main subject/topic of the review will be compared (e.g. another intervention or a non-exposed control group).
not relevant

22   Types of study to be included
Give details of the study designs to be included in the review. If there are no restrictions on the types of study design eligible for inclusion, this should be stated.
Inclusion criteria Studies will be eligible for review if they meet the following criteria: • economic evaluations • On PA promotion • For primary prevention purposes • Written in the English language Exclusion criteria Economic studies that are: • Not full economic evaluations (e.g. cost-analysis) • Not solely on PA promotion: o On programmes promoting PA in combination with other behaviours or technologies (e.g. dietary habits, smoking, drugs). • Not on primary prevention: o For health care purposes - (i.e. "curative" or "therapeutic", e.g. management/treatment of disease, rehabilitation) o For secondary prevention purposes (e.g. in diabetic or cardiac patients) o For prevention of not permanently debilitating, clinical conditions or symptoms (e.g. acute conditions, musculoskeletal disorders, pain).

23   Context
Give summary details of the setting and other relevant characteristics which help define the inclusion or exclusion criteria.

24   Primary outcome(s)
Give the most important outcomes.
Appraisal of the methods used for economic evaluation

Give information on timing and effect measures, as appropriate.

25   Secondary outcomes
List any additional outcomes that will be addressed. If there are no secondary outcomes enter None.
Narrative synthesis of findings of included studies

Give information on timing and effect measures, as appropriate.

26   Data extraction (selection and coding)
Give the procedure for selecting studies for the review and extracting data, including the number of researchers involved and how discrepancies will be resolved. List the data to be extracted.
Initial screening of titles and abstracts against inclusion criteria will be undertaken by one researcher (myself) and potentially relevant articles will be retrieved. A random 20 percent of the articles screened by title and abstract and all of the records assessed full text will be reviewed by a second researcher (DM). Any disagreements that arise between the two reviewers will be resolved through discussion.

27   Risk of bias (quality) assessment
State whether and how risk of bias will be assessed, how the quality of individual studies will be assessed, and whether and how this will influence the planned synthesis.
Quantification of bias effects in the studies will not be conducted as I do not intend to synthesise the cost-effectiveness results. In term of review, bias will be reduced by having two independent assessors (myself and DM)

UNIVERSITY *of York*
Centre for Reviews and Dissemination

NHS
**National Institute for
Health Research**

and by retrieving articles from a number of databases. There will be restriction to the published literature.

28   Strategy for data synthesis
Give the planned general approach to be used, for example whether the data to be used will be aggregate or at the level of individual participants, and whether a quantitative or narrative (descriptive) synthesis is planned. Where appropriate a brief outline of analytic approach should be given.
No meta-analysis is planned

29   Analysis of subgroups or subsets
Give any planned exploration of subgroups or subsets within the review. 'None planned' is a valid response if no subgroup analyses are planned.
None planned

Review general information

30   Type and method of review
Select the type of review and the review method from the drop down list.
Systematic review

Public health (including social determinants of health)

31   Language
Select the language(s) in which the review is being written and will be made available, from the drop down list. Use the control key to select more than one language.
English

Will a summary/abstract be made available in English?
Yes

32   Country
Select the country in which the review is being carried out from the drop down list. For multi-national collaborations select all the countries involved. Use the control key to select more than one country.
England

33   Other registration details
Give the name of any organisation where the systematic review title or protocol is registered together with any unique identification number assigned. If extracted data will be stored and made available through a repository such as the Systematic Review Data Repository (SRDR), details and a link should be included here.

34   Reference and/or URL for published protocol
Give the citation for the published protocol, if there is one.
Give the link to the published protocol, if there is one. This may be to an external site or to a protocol deposited with CRD in pdf format.
http://www.crd.york.ac.uk/PROSPEROFILES/52593_PROTOCOL_20161101.pdf

I give permission for this file to be made publicly available
Yes

35   Dissemination plans
Give brief details of plans for communicating essential messages from the review to the appropriate audiences.
As part of a PhD thesis, findings of the review will be presented at relevant national / international conferences and a paper will be submitted to a scientific journal for publication.

Do you intend to publish the review on completion?
Yes

36   Keywords
Give words or phrases that best describe the review. (One word per box, create a new box for each term)
economic evaluation, cost-effectiveness, physical activity

UNIVERSITY *of York*
Centre for Reviews and Dissemination

NHS
National Institute for
Health Research

37   Details of any existing review of the same topic by the same authors
Give details of earlier versions of the systematic review if an update of an existing review is being registered, including full bibliographic reference if possible.

38   Current review status
Review status should be updated when the review is completed and when it is published.
Ongoing

39   Any additional information
Provide any further information the review team consider relevant to the registration of the review.

40   Details of final report/publication(s)
This field should be left empty until details of the completed review are available.
Give the full citation for the final report or publication of the systematic review.
Give the URL where available.

Appendix M

SYSTEMATIC REVIEW SEARCH STRATEGY

1. (economic* adj evaluat*).tw.

2. (cost* adj (effect* or util* or benefit or consequenc* or minim*)).tw.

3. Cost-Benefit Analysis/

4. 1 or 2 or 3

5. Models, Econometric/ or Models, Economic/

6. Markov Chains/

7. Decision Trees/

8. Decision Support Techniques/

9. microsimulat*.tw.

10. (patient level adj simulat*).tw.

11. (simulat* adj model*).tw. and decision*.mp.

12. (discrete event* adj simulat*).tw.

13. (discrete event* adj model*).tw.

14. (decision adj model*).tw.

15. markov*.tw.

16. ((econom* or cost or costs) adj model*).tw.

17. "state transition model*".tw.

18. ("transition probabilit*" and (state or states or model*)).tw.

19. "health state*".tw.

20. ("disease state*" and (econom* or cost* or qaly* or utilit*)).tw.

21. or/5-18 [WITHOUT txt search for health state]

22. or/5-20 [WITH txt search for health state]

23. 4 or 22

24. Motor Activity/

25. exp Physical Fitness/

26. exp Sports/

27. Exercise Therapy/

28. exp Exercise/

29. (physical* adj2 activ*).tw.

30. gym.tw.

31. (physical* adj2 exerc*).tw.

32. cycling.tw.

33. walk*.tw.

34. danc*.tw.

35. jog*.tw.

36. (aerobic* adj exerc*).tw.

37. bicycl*.tw.

38. swimming.tw.

39. (fitness adj5 exerci*).tw.

40. (aerobic* adj5 fitness).tw.

41. (physical* adj5 fit*).tw.

42. sport*.tw.

43. or/24-42

44. Life Style/

45. lifestyle*.tw.

46. Attitude to Health/ or Health Behavior/ or Health Promotion/

47. (health adj prevent*).tw.

48. ((health* or ?activ* or change* or intervent*) adj3 behavio?r*).tw.

49. (promot* adj3 (health or physical activity)).tw.

50. sedentar*.tw.

51. (physical adj inactiv*).tw.

52. habit*.tw.

53. (physical* adj2 ?activ* adj3 (minute* or level* or participation or attendance or recommend* or proportion)).tw.

54. or/44-53

55. 23 and 43 and 54

56. limit 55 to (english language and humans)

Appendix N Systematic review of modelling studies manuscript

Working towards a consensus on how to model the impact of physical activity interventions on public health

Background

The finite resources available to decision makers dictates that commissioning of interventions ought to be based not only on effectiveness, but also on cost-effectiveness grounds REF. To support decision making concerned with funding interventions where there are multiple options, economic evaluation (EE) is typically used.

Such reimbursement decisions for health technologies, such as drugs and medical devices, requires a formal assessment adhering to established quality standards and agreed practices REF. For technology appraisals and public health, the National Institute for Health and Care Excellence (NICE) requires the use of methods of economic evaluation that adhere to the reference case REF. For public health however this represents a challenge. Public health covers a very broad range of topics, from disease prevention to health promotion, and unlike for health technology assessments, the reference case represents only a general guidance rather than a rule REF. As a result, key choices regarding the methods of economic evaluation are left to the discretion of the individual researchers. This is especially problematic considering the complexity of evaluating health promotion interventions REF, as being likely to result in wide variation in the structures and assumptions of the economic models, even within the same field, with implications for consistent and justifiable resource allocation decisions in public health. The promotion of physical activity (PA) in the general population is a priority for many public agencies across the world REF. Evidence demonstrates that physical inactivity increases the risk of many chronic diseases REF, determining 9% of all premature

mortality worldwide REF https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)61031-9/fulltext and having non-marginal impacts on national health care budgets REF. In the UK, physical inactivity costs around £1 billion a year to the national health system, with estimates rising to around £7.4 billion a year when taking a wider societal perspective REF. https://www.ncbi.nlm.nih.gov/pubmed/21562029.

What is less well-established, however, is how improvements in PA affect public health. While regular PA has been associated with risk reductions in many chronic diseases, the evidence for part of them is still limited or unclear REF. Furthermore, disease incidence and progression may vary significantly depending on personal characteristics (e.g. osteoporosis incidence in men vs women), be exclusive (e.g. prostate cancer in men), or be more or less relevant to certain groups depending on the time horizon considered (e.g. falls vs cancer in the elderly).

Moreover, different individuals may also respond heterogeneously in terms of PA behaviour change to the same level of intervention exposure. For instance, for sedentary individuals and from low socio-economic backgrounds, improvements in PA will be harder to achieve, relative to the non-sedentary and well-off. However, these can benefit the most from changes in PA, as being at a disproportionally higher health risk than the other groups REF.

In addition, changes in PA induced by interventions are also likely to be time-dependent (e.g. decays of effect over time), as well as differ, again, according to baseline characteristics and type of interervention. Large part of the generated health benefits are likely to occur after the observation period and when the active intervention has ceased. As a result, extrapolation over longer periods of time is typically needed REF.

Further, society values reducing existing unfair health inequalities between subgroups REF. Therefore, in order to assess the impact on public health, models need be able to take into account these differential effects (i.e. heterogeneity) not only for purpose of accurate population-level estimations, but to produce equity-relevant information REF.

A number of reviews have summarised the economic evidence for promotion of PA in the general population REF, finding the interventions to be cost-effective in the majority of cases REF. However, to date no review has assessed whether and how the complexities described above have been handled in practice in EEs. Such an assessment is important to guide future methodological research and work toward a consensus on minimum modelling standards.

Methods

Search strategy

A search concept tool as applied to structure the inclusion criteria. A method consisting of using some of suggested "PICOS" concepts (i.e. "I" for intervention and "S" for type of study") was chosen (Centre for Reviews and Dissemination, 2009). Free-text terms, synonyms, spelling variants, abbreviations and indexing terms (e.g. subject headings) related to three concepts were used: (1) EE, (2) economic model, (3) PA. No manual search using reference lists of existing literature was planned.

Validated search filters for identification of the relevant literature were not available. Search strings were developed from terms identified in known relevant publications and related to those three concepts. Concepts were combined using Boolean logic, as follows: (1) EE "OR" (2) economic model and the resulting (1+2) "AND" (3) PA. Other search filters, such as for intervention setting or type (e.g. community-based or workplace), were not

included as eligible papers could be missed. No limit to publication date or to the unpublished literature were set.

Eligibility criteria

Prospero database confirmed the absence of any ongoing reviews. Studies were eligible for inclusion if they met the following criteria:

- Type of study: any type of full EE. Partial EEs, such as cost-analyses were excluded.

- Intervention: any intervention aimed to promote PA behaviour (being either the focus of the study or one of the comparator interventions). Curative or rehabilitation programmes or studies evaluating the impact of hypothetical scenarios of changes in behavioural patterns (e.g. shift in number of active travellers) or associated health risks were excluded. As were those promoting PA in combination with other technologies or interventions (e.g. health dietary habits). Combined interventions cannot be fully comparable, because they address different yet closely related and multifaceted issues (e.g. obesity) and it can be particularly difficult to disentangle the combined effects on the economic results.

- Population: non-clinical populations. EEs whose study populations were targeted or selected on the basis of pre-existing disease conditions were not included (i.e. disabled individuals or secondary interventions in cardiac patients). Studies targeting "high risk" individuals, that is, clinically stable but carrying medically relevant conditions, such as hypertension or mild/moderate depression were included.

- Written in the English language (to allow for cross-checking).

Study screening and selection

 Identification of relevant articles was performed by screening against inclusion and

exclusion criteria. If there was insufficient information in the retrieved article, the

corresponding author/s were contacted to obtain the full text. After removal of duplicates,

initial screening of titles against inclusion criteria was undertaken. This step resulted in a

number of records to screen by abstract, with excluded references that were grouped in

relevant categories. Screening of abstracts followed, excluding articles on the basis of

study type (i.e. not full EE), intervention type (i.e. not solely on PA promotion) and target

population (e.g. cardiac patients). Following a procedure comparable to that followed in the

review by Alayli-Goebbels et al. (2014), a random 20 percent of the articles screened by

title and abstract and all of the records assessed full text were reviewed by a second

researcher. Any disagreement was resolved through discussion.

Data extraction

Data extraction forms were developed by adapting existing templates suggested by review

guides (Centre for Reviews and Dissemination, 2009, Joanna Briggs Institute, 2017) and

in a review by Weatherly et al. (2009). These forms were designed to capture contextual

and key methodological elements relevant to the set objectives (Centre for Reviews and

Dissemination, 2009). For all studies, only the information presented in the original

publication was used.

Assessments

As recommended for methodological reviews REF, assessments were provided in the

form of narrative summaries. An overview of the modelling approaches was first given. A

number of mathematical / statistical frameworks can be used to represent the PA – health

improvements processes, at different levels of sophistication and with different advantages

and disadvantages REF. Building on previous taxonomies developed by Brennan and

Squires, Briggs et al. (2016 https://pophealthmetrics.biomedcentral.com/track/pdf/10.1186/s12963-016-0085-1) have categorised public health economic modelling approaches based on

whether they are population (aggregate) or individual-level, time is formally modelled and their ability to capture interactions between the modelled entities and the environment REF. Using this classification, a description of the models was given also including details on the decision contexts, the downstream disease risks, as well as the final endpoints considered. The second part of the review focussed on a critique of the elements and structural assumptions relating to the complexities described in the background section, namely:

- Reflecting heterogeneity

- Modelling the mechanics of change in PA;

Results

Figure 1 shows a flow chart of the review stages. Twenty-five papers met the selection criteria, which included 26 modelling studies. Eleven papers based their analyses on primary data from the United Kingdom, seven from the US, four from Australia and one from Belgium, Canada and the Netherlands, each. The majority of studies focused on adults (>=18 years, n=20), four analyses focused on school pupils and two also included populations of children (<18 years).

Modelling approaches

Table 1 provides an overview of the reviewed studies. Eighty-one percent of the studies (n=21) employed aggregate-level approaches, nine of which used untimed modelling methods (eight comparative risk assessments and one decision-tree). Twelve analyses were based on discrete-time frameworks, with two multiple cohort life-table approaches (Cobiac,Zapata) and Markov chain modelling being used the most frequently (n=10). Of the five identified individual-level models, two were Markov chains (Goyder, Nshimyumukiza), one applied a system of linear equations using a cross-sectional regression analysis approach (Guo and Gandavarapu 2010), one a microsimulation

approach (Cradock, although no details were reported in terms of Markovian assumptions or interaction-levels) and one study developed a discrete event simulation model (Singh). Three studies used freely available off-the-shelf tools to conduct their evaluations (Cope, Dallat, Montes).

Modelling of downstream disease risk

The majority of studies (n=23) evaluated the impact of interventions on chronic diseases and conditions associated with PA, with eight of these studies not stating which diseases were considered. The number of chronic diseases selected ranged from one to seven, with one paper modelling 32 disease combinations REF. Type II Diabetes and at least one cardiovascular disease (either a type of Stroke or Coronary Heart Disease) were selected in all but one (n=14) of the papers reporting relevant details (n=14), which focused on Osteoporosis outcomes only. Eleven models included at least one cancer (i.e. Colon, Colorectal, Breast, Lung and Kidney) and only two studies considered impacts on mental health outcomes, specifically, depression. Only Munro included exercise-related injuries among the consequences. Within studies focussing on adults from the general populations (n=14), the majority (n=8) selected five chronic conditions. Choice of disease matched in three models that selected three diseases and three models that identified five diseases.

Final endpoints

Six studies considered impacts of intervention only on one health outcome. Cobiac considered changes in mortality risks, one of the two modelling studies by Beale modelled changes in QALY and four studies assessed the impact on changes in healthcare costs associated with changes in PA levels. Eighteen models considered impacts on health care

costs, as well as on generic measures of health. The majority of these studies (14/18) considered QALY gains, while four and one studies used DALY and HALY as primary outcomes, respectively. The remaining three studies, all of which focused on school pupils, considered obesity outcomes as final endpoints.

Reflecting heterogeneity

Table 2 provides a summary of how the issues of heterogeneity and modelling the mechanics of change in PA have been handled in practice in the reviewed studies. Ten studies used simple average approaches, evaluating the health impact of changes in PA levels in homogeneous groups of inactive / sedentary adults REF or school pupils REF. Baseline differences in PA were taken into account in only nine studies. From three to five levels (i.e. PA states) were defined in these models, with the models by Frew, Over, Pringle, 2x Roux and Zapata aligning the classification of PA levels to current national-level PA recommendations. Eleven studies accounted for heterogeneous health impacts based on at least age or gender, with Guo and gandavarapu + Singh also considering ethnicity/race differences. Health equity concerns were not formally incorporated in any of the reviewed economic models.

Modelling the mechanics of change in physical activity

Based on what was reported in the full papers, except for four studies, the large majority of models assumed that changes in PA would correspond to immediate gains in health outcomes. Anoyke assumed that the intervention could not affect disease risk in the first year ("run-in period"). Barrett assumed that it would take two years, while Cope et al five years, for the intervention to reach full effect, respectively. Except for Dallat, who reported

on the time lags between changes in PA and disease occurrence used in the model, none of the other studies addressed this aspect formally.

Except for the Markov model used in the studies by Roux, none of the reviewed models accounted for natural fluctuations in PA levels over time. PA states were assumed to be stable and with transitions between the highest and the lowest levels not being allowed. All other model did not model negative intervention effects, e.g. due to injuries or current exercisers put off by the intervention.

The majority of evaluations (n=15) considered time horizons equal or longer than 30 years, with all the studies that employed untimed modelling approaches (n=9) considering time horizons equal of longer than 10 years for their economic evaluations. The majority of models (15/26) assumed implicitly or explicitly that the intervention effect would not decay after the intervention ended (i.e. beyond follow-up assessment period). The remaining 11 analyses assumed a constant and homogeneous decline in effect, ranging from 25% to 100%, up to two years after the intervention ended.

Discussion
This review examined the modelling approaches used in previous economic evaluations for determining health impacts of changes in PA in the general population. Overall there is poor quality of reporting, which hindered the review process. Key structural assumptions regarding decay of effectiveness over time, dose-response relationship between changes in PA and health improvements were not made explicit in the majority of cases, making assessment of the modelling studies and interpretations of their results difficult.

Appendix O

A systematic review of economic evaluations of physical activity promotion interventions

Background

Review questions

Considering the two-fold aim of a comprehensive overview and an in-depth analysis, the questions posed related to two distinct parts of the review were as following:

Phase 1 (overview):

- What is the existing EE evidence base of interventions aimed to promote PA in primary prevention/non-clinical populations

Phase 2 (in-depth analysis of a sub-collection of phase 1 included EEs):

- Which and how appropriately have analytic methods been applied for EE of PA interventions designed to encourage participation in sport and exercise through provision of convenient access (in terms of proximity and/or membership cost) to leisure centre-based programmes / facilities (hereinafter referred to as "leisure centre-based interventions")?
- To what extent are the findings of EEs valid and applicable to the current UK PH decision-making context?

Methods of review

As mentioned above, the search strategy developed for the meta-review (please refers to sub-sections 3.2.1.1 and 3.2.1.2) mirrored in large part that used to retrieve the primary EE studies. Thus, in order to avoid repetition, only the pieces not in common with those used for the scoping exercise are included in the present section. However, the remaining review methods, namely, inclusion and exclusion criteria, study screening and selection, data extraction, quality assessment and data synthesis and reporting are described in the following paragraphs.

Search strategy

Like for the scoping exercise, the broadness of the questions formulated by the review was reflected into the four broad concepts used to identify relevant papers. Namely, economic evaluation, economic model, physical activity and behaviour/lifestyle, and their related terms (see lists in appendix…). No search filter

was applied. Intending to be the first review focusing on assessment of the methods rather than the estimates of EEs of PA promotion interventions, the search results were not limited to publication date or grey literature. In addition, given that the search had the purpose of informing also other parts of the thesis, no filter to type of publication was used. This allowed for identification of a number of relevant references, which were classified by study type and content for future use.

Inclusion and exclusion criteria

Studies were eligible for inclusion if they met certain criteria, as follows:

- Type of study: any type of economic study.

- Intervention type: any type of intervention of PA promotion.

- Written in the English language (to allow for cross-checking)

As mentioned above, those broad criteria were established to gather the wider literature regarding the economics of PA promotion, which served in informing several parts of the thesis. However, considering the systematic review aim of examining EE evidence of PA promotion interventions, more strict criteria were established for excluding non-relevant references. With a narrower scope, thus, exclusion boundaries were set. In particular, references were excluded from the review if they did not meet the following requirements:

- *Type of study*: any type of full EE, as studies providing efficiency information for resource allocation decisions. Full EE, defined as an empirical study in which both the cost and consequences of comparative interventions are assessed for the purpose to address a defined decision problem (i.e. cost-consequences analysis, CCA; or cost-effectiveness analysis, CEA; or cost-utility analysis, CUA; or cost-benefit analysis, CBA). Partial types of EE, such as cost-minimization analyses, cost or outcome descriptions, cost analyses, cost-outcome descriptions were not included ref 20 ghislaine. As were those evaluating the impact of hypothetical scenarios of changes in PA behaviours or associated health risks (e.g. health impact assessments of hypothetical change in the number of active commuters) as not considering intervention options or scenarios.

- *Intervention type*: any intervention aimed solely to promote increase and/or maintenance in physically active behavioural patterns (i.e. occupational, leisure-time, transport, home-based). Physical activity interventions are often part of multifaceted programmes, for example, PA promoted in combination with healthy dietary habits. However, these interventions cannot be fully compared with those aimed at promoting PA behaviours, as it can be particularly difficult to disentangle the combined effects on the economic results and because they address different yet closely related research questions (e.g. obesity prevention). Given the review scope and focus on PA behaviour change initiatives, composite interventions were thus excluded. As were those testing the cost-effectiveness of technologies or programmes specifically designed to improve physical fitness, rather than to change behaviour, in vulnerable groups of participants (e.g. fall or fracture

prevention in elderly)(is this a sensible and defensible distinction considering the purpose of this work?). Although these interventions may adopt similar intervention approaches to those aimed at encouraging sport and exercise participation (e.g. community-based exercise programmes), the addressed research questions do not overlap. In fact, the outcomes of interest within EE are different (e.g. changes of baseline measures of physical strength or resistance), yet closely related, as improvements in physical fitness or function can be health-intermediate outcomes of changes in PA behaviour or lifestyles.

- *Study purpose /population*: primary prevention/non-clinical populations (including of healthy, apparently healthy, at increased lifestyle risk, at increased disease risk groups of individuals). Interventions focused on physically impaired individuals (e.g. disabled), on patients already diagnosed with any chronic non-communicable disease or conditions (CNCD, e.g. cardiovascular disease, type 2 diabetes), or aimed to prevent or manage acute clinical conditions (e.g. back pain, curative or rehabilitation programmes) were excluded.

No other criteria, such as type of comparators, outcomes, source of effectiveness evidence, and duration of intervention or follow-up period were specified to restrict inclusion. However, relevant details about the included studies are reported in the following Results section.

Study screening and selection

Unlike for the scoping review, identification of relevant articles was performed by screening against inclusion and exclusion criteria. However, the base of references from which to start screening for relevant articles was the same as that gathered for the scoping exercise (please see paragraph 3.2.1.3). If there was insufficient information in the retrieved article, the corresponding author/s were contacted to obtain the full text. After removal of duplicates, initial screening of titles against inclusion criteria was undertaken. This step resulted in a number of records to screen by abstract, with excluded references that were grouped in relevant categories. Screening of abstracts followed, excluding articles on the basis of study type (i.e. not full EE), intervention type (i.e. not solely on PA promotion) and purpose/population (i.e. primary prevention, that is, in the general population). This framework was also used to classify full text papers that were not included in the review. Given the intention to submit for publication in a scientific journal, cross-checking was planned. Following a procedure comparable to that followed in other reviews ref Goebbels https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3299641/, a random 20 percent of the articles screened by title and abstract and all of the records assessed full text were reviewed by a second researcher (DM). Any disagreement was resolved through discussion during supervision meetings, thus a third reviewer was not necessary.

Data extraction

Standardised forms developed adapting existing templates suggested by review guides (York, brigggs) and papers (weatherly) were used to guide the data extraction process. These forms were designed to capture key methodological elements relevant to the posed review questions ref York guidance. For all studies, only the information present in the original publication was used. After several revisions and feedback from

supervisors, data extraction forms were defined and are available for consultation in appendix ….. In order to answer the first question, the following data were extracted from all the included EEs:

- Year of publication
- Country of origin / investigated health care system
- Promotion level/ approach/setting
- Target population
- Policy category
- Type of intervention / outcomes
- Target diseases/s
- Type of study / framework
- Economic findings
- Uncertainty assessments

Quality assessment

With regard to the first phase of the review, an illustration of the quantity and type of existing literature in relation to key contextual and methodological information concerning the studies was provided. More specifically, an overview along with an overall appraisal of the evaluation approaches used within the included studies was performed. These results also allowed for obtaining an indication of in which areas, within the investigated topic, EE evidence was scarce or even absent. (I'd like to discuss whether I should or not conduct this comparative analysis)

As for phase two….(I haven't found any framework for structuring the informing of the planned case study EE and decision model from existing studies)

Data synthesis and reporting

Results of review

The systematic search yielded a total of 6951 records. After removal of duplicates, articles were screened by title. The majority of articles were discarded at this initial stage as lacking of minimum requirements for inclusion (e.g. non-economic studies). After screening the abstracts, 54 full texts were selected for retrieval. Two articles referring to primary papers published by different authors (Cavill 2011, Patrella 2006) were retrieved in full text and included, while 19 were dropped as failing to meet exclusion criteria (e.g. partial EEs). Thirty-five unique articles fulfilling the inclusion and exclusion criteria were thus retained for review. A PRISMA-style diagram depicting the flow of information through the different phases of identification, screening and selection is displayed in Figure….

Characteristics of the included studies

In order to answer research question n. 1, that is, to address phase 1 of the review, the included studies were grouped by relevant categories to allow for discussion of the main characteristics. They are summarized in Table n……, to which I refer the reader for category-specific details. An expected degree of disparity in the methods used was found across the EEs, making it difficult to synthetise them into a coherent whole. However, an overview of the included EEs is provided in the following paragraphs.

Phase 1

The majority of studies (19) was published in the last six years, confirming a marked upward trend in the number of EEs performed on the research area, which more than doubled comparing with the previous two decades (1990-2010). Nevertheless, this growth in economic research was not spread evenly across countries. Almost three quarters (25/35) of the empirical investigations were conducted in or concerned the health systems of the UK or the USA. Four studies were carried out in the Australian continent and continental Europe contexts (The Netherlands, Spain and Belgium), respectively, only one in Asia (Taiwan) and the remaining project across four American countries (Mexico, Colombia and California, Montes). In what follows, an outline of the main features of the studies is given using PICOS concepts to help frame their description.

P – Population

Starting with the target populations, post-hoc classifications could not suit neatly, as evaluations often covered multiple age groups (e.g. adults and older adults)or defined inconsistently. Or, because authors did not document the age of participants (two cases, Montes and Wang). However, adults were subject of EE in the large majority of studies (28/35). Five EEs focused on older adults (defined as being at least 60 years old) and 5 on young people (children and adolescents). in addition, it is worth noting that in 16 of the 35 studies, the economic sample did not coincide with that of the effectiveness source (e.g. trial). This was the case of EEs designed to assess the health economic impact of defined alternative courses of action on hypothetical cohorts or entire populations inferring from effectiveness study samples (i.e. applying modelling techniques).

I – Intervention

In 20 of the 35 studies, population level interventions were considered for economic analysis, whereas seven and two studies adopted individual level or a combination of promotion approaches, respectively. This classification method is not universally agreed, yet is widely adopted as a way to distinguish between promotion approaches ref 13 michie BCW. However, population level interventions are usually characterized by wide reach and, unlike individual level ones, there is no active identification of potential participants (typically, a health care professional recommending or prescribing "high risk individuals" to take part into exercise schemes). In population level interventions, the promotion is carried out towards individual subjects as they belong to wider target groups or communities.

With regard to the level of promotion, within the group of population level interventions, a distinction was made between universal and targeted strategies. The difference between those attributes was defined as whether the intervention was made available to everybody within the identified group or community (universal strategies) or to only those individuals targeted as being (more) in need (targeted strategies). Across the included studies, certain socio-demographic and personal characteristics were used to identify potentially (more) in need individuals within communities, as these are generally associated with higher health risks in the relevant literature. Namely, age, gender, socio economic status (economic deprivation), ethnicity (minorities), lifestyle-related (e.g. sedentary job) and clinical conditions (e.g. increased blood

pressure, cholesterol or impaired glucose tolerance) were used. Within the included EEs, there appeared a balance in the number of studies per respective type of strategy, with just over half (19) adopting universal approaches.

Finally before discussing the interventions, in 20 studies the promotion of PA was carried out in community settings , whereas eight and five EEs considered initiatives promoted in primary care (e.g. GP practice) and occupational settings (four in schools and one at the workplace). The remaining two considered multiple interventions implemented in more than one setting.

Almost a third of the studies could not be classified within the behavior change policy framework (Michie et al). They were multi-component interventions or combinations of different intervention modalities (refs…roux and cobiac?). Ten studies were grouped as belonging to "Communication" or "Marketing" types of policies, as including interventions primarily based on providing health advice, counselling, media campaigns or written information. Four papers were categorized as "Environmental". These assessed the cost-effectiveness of changes in the built environment, such as building side walks, multi-use trails, cycle infrastructures and urban regeneration projects. Finally, the remaining nine studies fit under the broad umbrella of "Service provision". These interventions were so classed as encouraging physically active behaviours through provision of PA opportunities in the form of, for example, convenient access to leisure centres, fitness programmes or active travel initiatives. For more details about references, intervention designs and components I refer the reader to Table n. ….

C – COMPARATOR

As indicated in the Methods section, no limit was set in terms of type of comparator for including a study. Briefly, across the included EEs, 50 implemented interventions or intervention scenarios (i.e. in prospective EEs) were evaluated against one or more control conditions. The latter was no intervention or current practice scenarios in 28 cases, which were found more often as implicit or ill-defined rather than explicit alternatives.

O – OUTCOME

Having set no restriction to the type of outcome or study and consequently type of consequences considered for EE, a plethora of effects/benefits/outcomes/consequences was found across the reviewed EEs. In order to simplify the reading of results and better describing the details regarding the outcomes of EEs, a main distinction was made. Studies were distinguished between those considering and /or valuing CNCD -related consequences (e.g. morbidity or related healthcare cost-savings) and those comparing alternatives in terms of relative changes in PA effects or intermediate outcomes. In respect to this distinction, the sample of studies was split unevenly, with 26 of the 35 belonging to the first group (disease-related consequences), of which 14 studies considered a combination of CNCDs. In particular, cardiovascular disease was the most prevalent CNCD, followed by type II Diabetes and certain types of cancer (in particular, breast and colorectal cancers). Only four studies considered explicitly mental health-related problems (e.g. depression) in the economic models. However, measurement of change in these medical conditions could have implicitly been included within estimations of changes in health risks in participants, by those studies employing non-disease related quality of life measures (e.g. QALY or DALY gain estimates in the general population). The second group of studies compared alternative interventions in terms of: PA effects (5 studies, e.g. minutes of PA, number of people becoming active) or intermediate outcomes (4 studies, e.g. METs). Clearly, the metrics used for comparing alternatives with one another (e.g. incremental ratios) depended on the type/s of EE framework used in the study, on which details are provided in the following paragraph.

S – STUDY

The included EEs were also categorized in respect to two methodological aspects: measurement / estimation of effects/outcomes and type of EE framework used. As for the first aspect, the sample of included studies was split into two groups, according to whether or not modelling techniques or methods (e.g. mathematical applications or DAM) were applied within a study. Each of these main groups were further divided in two sub-groups. The majority of the EEs (14) was based on single study-based estimates (i.e. 12 RCT, one controlled trial and one observational study), while one only (ref Babey 2014) on synthesis-based estimates (i.e. meta analyses). Within the group of studies applying some form of modelling (e.g. to infer to hypothetical populations or to link changes in PA effects with future health benefits), ten studies were purely model-based, that is, they based their analyses on hypothetical populations with primary data inputs (e.g. effectiveness data) collected from secondary sources (e.g. meta-analysis or literature reviews). Only a fifth of the sample of studies applied "mixed-method" approaches. Specifically, for these economic analyses the source of effectiveness was a single study (experiments / quasi-experiments in six EEs and one observational design) with modelling techniques applied mostly (in six of seven cases) to estimate long-term gains in health-related quality of life.

It is also worth noting that for all EEs, except one (ref Babey 2014), incremental rather than average analyses of costs and consequences were performed, although these definitions obviously coincide in those evaluations comparing only one alternative to a "doing nothing" option. In addition, all evaluated alternative courses of action were assesses as independent options.Finally describing the types EE frameworks used, according to the set inclusion criteria any type of full EEs could be retained for review. Overall, full EE frameworks were applied 40 times within the 35 included studies. Two thirds of all frameworks used were CEA or CUA, with the latter being the most prevalent approach employed for EE (19). CBA was applied singly in four studies, all including environmental types of intervention, and in two studies in combination with CUA. CCA was only performed in the less recent study and in combination with CEA (Murno 1997). Eight of the 35 studies employed multiple frameworks (all of them used two EE frameworks) for comparative analysis of the considered alternatives. More details are available in Table n…

Economic findings

Narrative summary…

Uncertainty assessments

Appendix P

MODEL PARAMETERS SEARCH TERMS

1. (risk*).tw.

2. (diabet*).tw.

3. (conorary*).tw

4. (heart*).tw

5. (stroke).tw.

6. (colo*).tw.

7. (bowel).tw.

8. (breast).tw.

9. (cancer).tw.

10. (frailty).tw.

11. (depriv*).tw.

12. (gradient).tw.

13. (socio-econon*).tw

14. (cost*).tw