Improving the validity and usability of decision models: case studies with a focus on physical activity

Robert Smith

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Medicine, Dentistry and Health

School of Health and Related Research

February 2023

Acknowledgements

I would like to thank my supervisors Professor Elizabeth Goyder, Dr Hazel Squires and Dr Chloe Thomas for their support and feedback throughout the PhD programme. In particular I appreciate their patience during a prolonged period of uncertainty throughout the pandemic, and their encouragement of my external collaborations. I would also like to thank Professor Alan Brennan who has provided mentorship and encouragement (often in A3 format) throughout my time at Sheffield. Finally, I would like to thank the directors and members of the Wellcome Trust Doctoral Training Centre who built an amazing programme which has supported my development throughout my PhD programme. I'm especially grateful to Dr Sarah Bates, Dr Tom Bayley, Genevieve David, Dr Simon McNamara, Wael Mohammed, and in particular Dr Paul Schneider for the interesting debates and productive collaborations from which I learnt a huge amount.

I am also grateful to colleagues from the various organisations with whom I have collaborated with; Professor Sonja Kahlmeier and Dr Thomas Gotschi from the WHO's HEAT team, Dr Fergus Cumming and Timothy Laurence from the UK Health Security Agency, Dr Rose Hart and Dr Darren Burns from BresMed, and Professor Steve Haake and Dr Helen Quirk from the parkrun research board. Also, thanks to those who provided their expertise to help shape the conceptual model for PACEM including; Dr Nick Cavill, Dr Mohammed Farooq, Dr Vijay Gc, Dr Emma Hock, Dr Oliver Mytton, Prof Harry Rutter, Dr Rebecca Robinson and Keith Tolley.

I'd like to thank my parents and friends who continually berated, mocked and ridiculed me until I finished the thesis. Most of all, I want to recognise the contribution of my wife Laura whose support was unwavering despite my unreasonably long working hours during the pandemic. She will always be the 'real' doctor in the house. Finally, for Isaac (age 2 years), and Oliver (age 2 days) who made things so much more difficult :)

Abstract

Background: Health economic evaluation has a crucial role to play in the allocation of scarce societal resources. Economic models used in these evaluations must have a high degree of external validity but must also be usable in order to effectively inform policy. However, there is sometimes a trade-off between the realism of models (external validity) and the ease with which stakeholders can understand and interact with them (usability). This trade-off is particularly relevant in the field of physical activity where modelling is complicated and data availability is limited. The aim of this thesis is to investigate the balance between the external-validity and usability of models used in health economic evaluations of physical activity interventions and develop ways to build models that are more externally valid and usable.

Methods: The study begins by identifying limitations in the external-validity and usability of published physical activity models, with a particular focus on models used to inform National Institute for Health and Care Excellence (NICE) guidance. Three case studies of adaptations to improve external validity are provided, with a discussion of their implications for usability. Additionally, ways to improve the usability of models are examined, with methods proposed to make models more accessible, transparent, secure, and efficient to construct and maintain.

Results: The results of this thesis demonstrate that models can be improved in terms of both external-validity and/or usability. The case studies provided show that methodological developments to physical activity models are feasible given new modelling methods and advancements in computing power, but despite improving external validity may reduce usability. Additionally, this thesis outlines methods by which health economic models can be made more accessible, transparent, secure, and efficient to construct and maintain, thereby improving their usability.

Discussion: The overall conclusion of this thesis is that economic evaluation models should be as externally valid and usable as possible. However, a trade-off sometimes exists between the two. With a fixed budget for evaluation, attempts to improve external validity can have an opportunity cost in terms of resources allocated to making models easy to use and understand. The incorporation of methods from computing and data science can help mitigate this trade-off.

Overview of Thesis

This thesis investigates the balance between the realism of models used for health economic evaluations (external validity) and the ease with which stakeholders can understand and interact with them (usability). The content begins by identifying limitations of published physical activity models and provides three case studies of adaptations to improve external-validity. It then examines ways to improve the usability of models in public health and health economics through the incorporation of modern data science techniques that make the models more accessible, transparent, secure, and efficient to construct and maintain. The aim throughout is to investigate the balance between the external-validity and usability of models used in health economic evaluations of physical activity interventions and develop ways to improve their external validity and usability to maximise the public health benefits of health economic evaluations.

This thesis is split into several parts:

Part 1 Defining the problem

Part 1 provides the context for the thesis and reviews the existing literature to identify how the cost-effectiveness of physical activity interventions has been estimated in the past. It contains two chapters:

Chapter 1. Background. This chapter introduces the topic, defining physical activity and why it is a public health issue, identifies interventions from the literature and discusses the need for the application of health economic methods to aid decision makers.

Chapter 2. Scoping review. This chapter reviews the literature to identify health economic analyses of physical activity interventions, trying to understand how physical activity has previously been modelled? What are the limitations of the current methods and what possible improvements could be made to improve the external validity and usability of models and therefore better inform decision makers?

Part 2 Adapting an existing model - case study of the WHO HEAT model

Part 2 describes the effect of two adaptations to a widely used tool, developed on behalf of the World Health Organisation, which aims to estimate the benefits of walking and cycling. The adaptations aim to improve model validity. Part 2 contains three chapters:

Chapter 3. The Value of the Statistical Life Year. A comparison of the HEAT walking and cycling model physical activity module results using the Value of a Statistical Life and Value of a Statistical Life Year methodologies.

Chapter 4. Incorporating the Dose Response Function into HEAT. A comparison of the HEAT walking and cycling model when using a linear and non-linear dose response function.

Chapter 5. HEAT Model - Discussion. Discussion of the differences using the Value of a Statistical Life Year (vs VSL) and non-linear dose response functions (vs linear DRF) and trade-off between model usability and external validity.

Part 3 Developing a new model: case study in modelling the long run cost effectiveness effect of school based physical activity interventions.

In part three I describe the methods used, and results of, a bespoke microsimulation model which estimates the cost-effectiveness of interventions which aim to increase childhood physical activity. The model was built to improve the external validity of physical activity models, but also to better understand the impact on model results of some of the structural assumptions identified as limitations in Chapter 2.

Chapter 6. PACEM model - Chapter 6 describes the methodology used to create a microsimulation model of Health-Related Quality of Life and healthcare costs. The results of the probabilistic model are reported with sensitivity analysis and a discussion of the significance of the findings for policy and comparison to other studies in the economic analysis of physical activity interventions in childhood.

Part 4 Improving usability, efficiency, and transparency of health economic models.

Part 4 contains a discussion of the use of web-based user interfaces for health economic models, demonstrating the value of these tools in improving the usability of models and explaining the method used to create such tools. It also describes the benefits of methods to semi-automate updates to health economic evaluations and provides a tutorial on how to achieve this using open-source software.

Chapter 7. On the creation of web application user interfaces for health economic and public health models. This chapter discusses the benefits of user-interfaces for health economic models, using a project undertaken with parkrun to highlight the benefits and limitations of these methods for public health economics. This chapter goes on to provide a tutorial on R-shiny for health economics more generally.

Chapter 8. Living HTA: Automating Health Economic Evaluation with R. This chapter provides a tutorial on automating updates to health economic evaluation reports using open-source software. The tutorial uses the teaching model as described in the previous chapter to show how the model can be automated to be re-run, and a new health economic evaluation report generated any time the underlying data used to populate the model is updated. Furthermore, it is demonstrated how this can be achieved *without* sensitive data being shared with the health economic modeller or any third parties.

Part 5 Concluding

The last part of the thesis discusses the trade-off between external validity and usability of models in the context of the work contained in parts 2-4, providing a conclusion to the thesis. It contains one chapter.

Chapter 9. Discussion and conclusion. This chapter starts with a discussion of the main findings of the thesis. It outlines the contribution of the work described in this thesis in identifying methods of improving the external validity and usability of health economic models, with a focus on models used to evaluate physical activity interventions. The chapter goes on to discuss the implication of these findings for policymakers and health economists. It then identifies the strengths and limitations of the thesis, before concluding.

Table showing summary of thesis structure

Chapter	Research Objectives	Intermediate Aim	Overarching Aim
Chapter 2 Review of limitations in existing PA models	Identify the limitations and key structural assumptions made in existing health economic evaluation models for physical activity	Improve the external validity of health economic models of physical activity interventions, while investigating the potential impact on usability.	Encourage the development of models that are the optimal combination externally valid and usable with a particular focus on health economic evaluations of physical activity interventions.
Chapter 3 Adaptation of the HEAT - Value of a Statistical Life Year	Understand the effect of adapting the HEAT to incorporate consideration of duration of life lost and weigh this against implications for model usability.		
Chapter 4 Adaptation of the HEAT - Dose Response Function	Understand the effect of adapting the HEAT to incorporate a non-linear dose-response function for mortality and weigh this against implications for model usability.		
Chapter 6 Development of PACEM - evaluating structural uncertainty	Understand the effect of adapting several structural assumptions and test the feasibility of a microsimulation model which models long-term PA trajectories.		
Chapter 7 Web-based user-interfaces for health economic models	Develop methods to allow health economists to build and deploy user-interfaces for script-based health economic models online to improve model usability and transparency.	Improve the usability of health economic decision models via new methods from data-science, with case studies and a consideration of the application of the methods to physical activity models.	
Chapter 8 Automating health economic evaluation model updates	Create a prototype open-source application to allow updates to Health Economic Evaluation reports to be automated as new information becomes available.		

Development of work over the course of the PhD

The methods and publications presented in this thesis were developed from September 2017 to May 2023. Through various projects, my perspectives evolved and influenced the work.

In 2018, I undertook a research visit with the WHO-Europe collaborating centre for physical activity at the University of Zurich. I worked with the Health Economic Assessment Tool modelling team to better understand the tool. Through this experience, I learned the significance of creating models that are easy to use and understand, even if it means sacrificing some level of external validity. This understanding greatly influenced the development of Chapter 3 and Chapter 4.

In 2019, I collaborated closely with a Wellcome Trust doctoral candidate, Paul Schneider, as well as academics from Sheffield Hallam University and staff from parkrunUK to inform decisions on the location of new parkrun events. Through this work, I gained a valuable understanding of the potential of these tools, in particular the use of web-apps to host models. Paul and I have since been asked to help build models with web-apps for several teams at the World Health Organization, who wanted to build and deploy models for a large number of geographical areas at once and required that these models be both transparent and usable by stakeholders worldwide. These experiences helped to inform and shape the development of Chapter 7 and to improve the content of Chapter 6.

Then the pandemic happened.

In 2020, I was approached by the Joint Biosecurity Centre (JBC) and invited to join their Advanced Analytics team to assist in informing the government's response to the COVID-19 pandemic. During my two years working with colleagues at the JBC (now United Kingdom Health Security Agency), I gained insight into the fast-paced nature of decision making in times of crisis, and the crucial importance of developing clear, easy-to-understand outputs. I saw first-hand the value of allowing decision-makers to interact with and experiment with models, and the tremendous impact it had on their decision-making abilities. I also developed an appreciation for the power of automating analysis and modelling updates to improve reporting efficiency, and the importance of data security. These experiences greatly influenced the development of Chapter 8.

Statement of Authorship

Chapters 3, 4, 7 and 8 contain papers or links to papers that have been published or are undergoing peer review. Chapter 3 contains a single paper that is published in *Public Health*, Chapter 4 contains a single paper that is under open peer review in *Wellcome Open Research*, Chapter 7 contains a paper published in *Wellcome Open Research*, and cites a body of work consisting of three papers published during my PhD, in *Wellcome Open Research*, *Health & Place* and *Public Health*. Chapter 8 contains a paper published in *Wellcome Open Research*. A summary of the authors' contribution, and their relevance, is included at the start of each chapter. The references for these papers are listed below:

Chapter 3

Smith, R., Thomas, C., Squires, H., Götschi, T., Kahlmeier, S., & Goyder, E. (2021). The price of precision: trade-offs between usability and validity in the World Health Organization Health Economic Assessment Tool for walking and cycling. *Public Health*, 194, 263-269. (https://doi.org/10.1016/j.puhe.2021.03.016)

Chapter 4

Smith, R., Thomas, C., Squires, H., & Goyder, E. (2022). A comparison of the World Health Organisation's HEAT model results using a non-linear physical activity dose response function with results from the existing tool. *Wellcome Open Research*, 7*:7 [version 2; peer review: 2 approved with reservations]* (https://doi.org/10.12688/wellcomeopenres.17411.2)

Chapter 7

Smith, R., & Schneider, P. (2020). Making health economic models Shiny: A tutorial. *Wellcome Open Research*, 5*:69 [version 2; peer review: 2 approved]* (https://doi.org/10.12688/wellcomeopenres.15807.2)

Chapter 8

Smith, R. A.**,** Schneider, P. P., & Mohammed, W. (2022). Living HTA: Automating Health Economic Evaluation with R. *Wellcome Open Research*. 7:*194 [version 2; peer review: 2 approved]*. (https://doi.org/10.12688/wellcomeopenres.17933.2)

Other authored publications cited in this thesis

The publications below were undertaken during my PhD and are cited in this thesis:

Smith, R., Schneider, P., Bullas, A., Haake, S., Quirk, H., Cosulich, R., & Goyder, E. (2020). Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England. *Wellcome Open Research*, *5:9. [version 2; peer review: 3 approved]*. (https://doi.org/10.12688/wellcomeopenres.15657.2)

Schneider, P. P., **Smith, R.** A., Bullas, A. M., Quirk, H., Bayley, T., Haake, S. J., ... & Goyder, E. (2020). Multiple deprivation and geographic distance to community physical activity events—Achieving equitable access to parkrun in England. *Public health,* 189, 48-53. (https://doi.org/10.1016/j.puhe.2020.09.002)

Smith, R. A., Schneider, P. P., Cosulich, R., Quirk, H., Bullas, A. M., Haake, S. J., ... & Goyder, E. (2021). Socioeconomic inequalities in distance to and participation in a community-based running and walking activity: a longitudinal ecological study of parkrun 2010 to 2019. *Health & place*, 71, 102626. (https://doi.org/10.1016/j.puhe.2021.03.016)

Haake, S., Heller, B., Schneider, P., **Smith, R.**, & Green, G. (2022). The influence of neighbourhood equity on parkrunners in a British city. *Health promotion international*, 37(2), daab138.

Thokala, P., Srivastava, T., **Smith, R.**, Ren, S., Whittington, M., Elvidge, J., Wong, R. & Uttley, L. (2023). Living Health Technology Assessment: Issues, Challenges and Opportunities. *PharmacoEconomics.*

Other publications not cited in this thesis but undertaken during my PhD

I was also fortunate enough to be involved in several other projects during my time undertaking a PhD at ScHARR. These publications helped shape my thinking but are not cited directly in the thesis.

Tordrup, D., **Smith, R.**, Kamenov, K., Bertram, M. Y., Green, N., & Chadha, S. (2022). Global return on investment and cost-effectiveness of WHO's HEAR interventions for hearing loss: a modelling study. *The Lancet Global Health*, 10(1), e52-e62.

Hounsome, L., Herr, D., Bryant, R., **Smith, R.**, Loman, L., Harris, J., ... & Youhan, U. (2022). Epidemiological impact of a large number of incorrect negative SARS-CoV-2 test results in South West England during September and October 2021. *medRxiv.*

Table of Contents

List of Tables and Figures

Abbreviations

BC	Breast Cancer
CC	Colorectal Cancer
EQ5D	EuroQol-5 Dimension
GBD	Global Burden of Disease
HEAT	Health Economic Assessment Tool
HRQoL	Health Related Quality of Life
HSE	Health Survey for England
HTA	Health Technology Assessment
ICER	Incremental Cost-effectiveness Ratio; also, Institute for Clinical and Economic Review (US)
IHD	Ischaemic heart disease
IS	Ischemic Stroke
JBC	Joint Biosecurity Centre
LSOA	Lower Layer Super Output Area
NICE	National Institute for Health and Care Excellence
PA	Physical Activity
PPF	Production Possibility Frontier
ScHARR	School of Health and Related Research
T2D	Type 2 Diabetes
UCL	University College London
UI	User Interface
UKHSA	UK Health Security Agency
UX	User Experience
VSL	Value of a Statistical Life
VSLY	Value of a Statistical Life Year
WHO	World Health Organisation

Part 1 - Introduction

This part of the thesis is made up of two chapters. Chapter 1 contains a short justification for the focus on physical activity (PA), defining what it is and discussing the effect of PA on different health and non-health outcomes and how risk from low levels of PA is accumulated. It then goes on to outline how PA is measured and common surveillance methods and what is happening to PA levels worldwide and in the United Kingdom (UK). Finally, it summarises the literature on the determinants of physical activity and government PA strategies and common interventions.

Chapter 2 includes a scoping review of the literature on the methods used in economic evaluations of interventions to increase population physical activity. It focuses on models submitted to the National Institute for Health and Care Excellence (NICE), before reviewing systematic reviews of other models contained in the peer reviewed literature.

Chapter 1. Introduction

Non-communicable diseases (NCDs) have risen in overall prevalence over the past century, such that in 2019 they are estimated to account for 74% of deaths worldwide and over 60% of deaths in every WHO region except Africa (Vos et al., 2020). The shift from communicable diseases to non-communicable diseases occurred first in developed countries but is well under way in developing countries (Yang et al. 2008).

Most NCDs have been shown to be strongly correlated if not causally linked with four risky behaviours: tobacco use, physical inactivity, unhealthy diet, and harmful use of alcohol (WHO, 2013). Insufficient physical activity is estimated to account for 21.5% of ischaemic heart disease, 11% of ischaemic stroke, 14% of diabetes, 16% of colon cancer and 10% of breast cancer as well as several other conditions worldwide (Bull *et al.*, 2004), amounting to around 2 million premature deaths annually (WHO, 2004). Increasing physical activity levels has the potential to reduce the prevalence of these conditions, thereby improving quality of life and reducing pressure on healthcare systems worldwide (Ding *et al.*, 2016).

Defining Physical Activity

Physical activity is defined by the World Health Organisation (WHO) as "*any bodily movement produced by skeletal muscles that requires energy expenditure*" (WHO, 2023). PA may take many forms, the NICE states that "*Physical activity includes everyday activity such as walking and cycling to get from A to B, work-related activity, housework, DIY and gardening. It also includes recreational activities such as working out in a gym, dancing, or playing active games, as well as organised and competitive sport*" (NICE, 2018). Exercise is a subset of PA, defined by Casperson *et al.* (1985) as "*physical activity that*

is planned, structured, repetitive, and purposive in the sense that improvement or maintenance of one or more components of physical fitness is an objective". Physical fitness is used to describe a set of attributes of an individual. It has multiple components; cardiovascular fitness differs from muscular strength, flexibility, body composition and bone density. There is ongoing debate as to the relative importance of PA and physical fitness (Blair *et al.*, 2001). However, PA correlates strongly with fitness and so there is agreement that physical activity is both a direct and indirect means of improving health (Ekelund *et al.*, 2007).

Since PA includes a variety of bodily movements, it became necessary to categorise different activities along a single dimension. Intensity provides that single dimension, allowing a gentle stroll to be distinguished from a vigorous swim, and a vigorous swim to be equated to a hard run. While there are other facets of PA, for example strengthening of the muscles and bones (Kohurt et al. 2004), and psychological routines (Penedo & Dahn, 2005), the bulk of the benefits accrue in cardiovascular and metabolic conditioning from aerobic activity (WHO, 2021) and can therefore be approximated by a measure of duration and intensity.

Duration is typically measured using minutes or hours, but there are numerous measures of intensity of PA, ranging from perceived exertion (e.g., the Borg Scale), % of maximal heart rate, accelerometers and more sophisticated techniques including doubly labelled water methods. Metabolic Equivalent of Task (MET) is becoming one of the most used by researchers focused on the epidemiology of physical activity related diseases (Ainsworth et al., 2000). A MET is a continuous unit of intensity of PA, measured relative to the basal metabolic rate (1 MET). Ainsworth *et al.* (2000) provide a compendium for adults, in which a typical healthy adult sitting using a computer is at approximately 1.5 METs, a normal walking pace is around 3 METs, a fast walk (4.5mi/hr) is 7 METs and running (6mi/hr) is around 10 METs. Similar values are available for children (Ridley *et al.* 2008). Weekly MET-mins or MET-hrs are easily calculated by multiplying intensity (METs) by duration (mins/hours).

This chapter has several aims, 1) to provide the reader with some background information around the benefits of physical activity and the health implications of low population levels of physical activity 2) to provide the reader with a sense of the scale of the problem, the extent to which the population is insufficiently active and the heterogeneity in activity levels between groups including ecological models of the determinants of physical activity, 3) to provide an overview of the role of the state in population physical activity levels, in surveillance, guidance and interventions.

Effect of Physical Activity on Health Outcomes

Physical activity has been linked to good health as far back as the philosophers of ancient Greece. A quote attributed to Plato (424–348 BC) states that a "*Lack of activity destroys the good condition of every human being, while movement and methodical physical exercise save it and preserve it*" (attributed to Plato). Despite this, the first epidemiological study of the effect of physical activity on

health is generally credited to Jeremy Morris in 1953 (Morris et al. 1953). Morris' study compared the rates of coronary heart disease (CHD) between drivers (inactive) and conductors (active) working for the London Transport Executive. Using data from around 50,000 person-years, the authors found that the more active conductors had a relative risk for CHD of approximately 0.7 compared to their inactive bus driving colleagues.

Since 1953 there has been substantial progress in physical activity epidemiology, to the extent that the literature is overwhelming, with many studies focused on different health outcomes, types of physical activity, on different populations, controlling for various confounding factors. Fortunately, several systematic reviews of the literature have previously been conducted to provide an indication of conditions for which the evidence base is strong. (Warburton et al., 2017; Kyu et al., 2016; Reiner et al., 2013). Table 1 below lists, in order of publishing date, the seven reviews identified by a non-exhaustive scoping review, and the conditions identified by the reviews as being strongly related to PA. Four of the reviews investigate the role of PA in all-cause mortality, all seven include diabetes, six cardiovascular diseases, three obesity, six breast cancer, six colon cancer and four mental health.

Table 1 Six Systematic Reviews of the Health Benefits of Physical Activity

Systematic review	ACM	T2D	CVD	Ob	BC	CC	MH	Other
Warburton *et al.* (2006)	X	X			X	X		Falls
PAGACR (2008)	X	X	X	X	X	X	X	Falls
Warburton et al. (2010)	X	X	X		X	X	X	Osteoporosis
Lee *et al.* (2013)		X	X		X	X		
Reiner *et al.* (2013)		X	X	X			X	
Warburton et al. (2017)	X	X	X	X	X	X	X	Gallstone Disease, SRH
Number	3	6	5	3	5	5	4	

SRH= Self-Reported Health; CVD = Cardiovascular diseases; T2D = Type 2 Diabetes; BC = Breast Cancer; CC = Colorectal Cancer; MH = mental health; Ob = Obesity; ACM = All-cause mortality

The most recent, published by Warburton et al. in September 2017, is a systematic review of systematic reviews of the relationship between PA and health outcomes. This is the most comprehensive summary of the evidence surrounding the benefits of physical activity. Within the 16 systematic reviews studied, the conditions found to be related to PA were: All-Cause Mortality, Diabetes T2, All-Cancer Mortality, Hypertension, Breast Cancer, Colon Cancer, Obesity, Gestational Diabetes, *Gallstone Disease*, Ischemic Heart Disease, Ischemic Stroke and potentially also *Self-Reported Health*. For most of the

aforementioned conditions there is a consensus in the literature that there exists a strong inverse dose response relationship between PA and risk (Davies et al., 2011). This relationship is thought to be curvilinear, with the greatest health benefits generated from increasing the physical activity levels of the least active (Kyu et al. 2016), although there is some disagreement as to the extent to which returns diminish (Bull et al. 2010, p.24). In fact, Warburton et al. suggest that the use of thresholds in public-health messaging is not supported by the literature, which suggests a clear dose-response relationship with no obvious threshold point.

It is important to recognise that while the benefits are large, there are also some health risks associated with physical activity, in particular cancer of the skin and musculoskeletal injuries. Moore *et al.* (2016) found that in their meta-analysis of 12 studies of the relationship between PA and malignant melanoma, including 12,438 cases, those who were at the 90th percentile of PA had a 27% increased risk relative to their inactive counterparts. The association was stronger in studies conducted in high UV areas, and in European subjects, suggesting that PA increases time spent outside and therefore sun exposure. This finding may be important for advising the public on when and where to exercise, and about the importance of adequate UV protection. Melzer *et al.* (2004) highlight several potential risks identified by the literature, including musculoskeletal injury, dehydration and heat stroke. They posit that the most common risks are musculoskeletal in nature, particularly of the knee, foot and back. While these injuries are not insignificant to those affected, the relative health and financial burden is limited in comparison to that of the non-communicable diseases described above. It is also unclear as to whether increasing PA at earlier ages increases or reduces the risk of PA over the life course (Janssen & LeBlanc, 2010) since increases in exposure may be mitigated with improvements in bone density and flexibility.

There is very little evidence of how the risk of the chronic diseases associated with lower PA levels are accumulated over the life-course. While some models of diseases have estimated trajectories of certain markers in relation to multiple risk factors (e.g. Breeze et al., 2016), a scoping review of the literature has not revealed any studies which analyse the relationship between PA level and risk over the life course for multiple diseases. The most basic studies simply include variables in a multivariable model which represents PA level at multiple points over the life-course, generally baseline and follow-up (Politt et al., 2005). Since the early 2000s some models have developed to consider critical periods within the life-course, such as childhood, and the modifiers of this effect, e.g. adulthood PA may attenuate the effect of childhood PA (Ben-Shlomo et al., 2002). Other epidemiological studies have aimed to better understand accumulation of risk, including risk clustering, which describes the clustering of risk factors such as smoking and alcohol use, and chains of risk in which a person has a higher probability of developing a new risk factor if they have one already, e.g. if a person becomes unemployed they may have a higher risk of inactivity, which results in higher later risks of diabetes or depression. Perhaps the most developed examples come from studies of inequalities, which measure socioeconomic status at different periods of the life-course, usually childhood and adulthood, to determine how childhood SES influences adult health, and whether the effect is attenuated by adult SES (Poulton et al., 2002). However, this approach is still very crude, and ideally a model would be generated in which risk of

diseases was a continuously measured variable, and risk increased or decreased in each period depending on health behaviours. Risk could increase every year a person was insufficiently active and decrease in later years if an individual was more active, and certain years could be weighted during certain parts of the life course to reflect critical periods. Creating these complicated models for PA would necessitate data from large and representative longitudinal studies which gathered information on PA, amongst other comorbidities, on a regular basis over the entire life course (Ben-Shlomo et al., 2002). It may be that over the coming decades the data from newer studies provide this necessary information, for example the Millennium Cohort Study (Griffiths et al., 2013) and the Physical Activity Longitudinal study (Craig et al., 2005).

Physical Activity Surveillance

Public health surveillance, as defined by WHO (2017), is the systematic collection, analysis and interpretation of health-related data to inform public health practice. In the case of non-communicable diseases, surveillance is used to monitor progress and allocate resources. Physical activity surveillance is a recent development and is used to monitor physical activity levels in the population. There is a significant gap in physical activity surveillance compared to other chronic disease risk factors (Bauman et al., 2009). However, as the prevalence of diseases for which physical activity is a risk factor has increased, national and international bodies have started to invest more resources in monitoring physical activity.

The UK has two main surveys that monitor physical activity levels: the Health Survey for England (HSE) and the Active Lives Survey (ALS). The HSE is an annual survey of 20,000 individuals in England, measuring lifestyle factors and health status, including some objective measurements by nurses (Craig et al. 2013). The 2015 version contains a detailed focus on physical activity in adults and children. The ALS, which took over from the Active People Survey (APS) in 2015 and surveys 200,000 individuals per year, focuses on adolescents and adults and provides more accurate geographical differences. However, data from the ALS is not directly comparable to either the HSE or the APS (2005-2015). The general physical activity question in the ALS asks about frequency, time spent, and effort level, but previous comparisons with the HSE's International Physical Activity Questionnaire (IPAQ) showed poor agreement (Zwolinsky *et al.*, 2015), making it difficult to monitor trends from periods prior to 2015.

Measurements of physical activity can be broadly categorised into subjective and objective. Subjective measurements rely on self-reported information, while objective measurements are collected using devices or technology. Subjective measurements, including questionnaires, activity logs, and diaries, have several benefits. They are cheap to administer, include all types of physical activity, and have good test-retest validity (Herrman et al. 2013) and therefore remain the predominant method used for physical activity surveillance. However, they also have limitations, including recall bias, interpretation of what constitutes moderate or vigorous activity, and overestimation of physical activity levels (Drystad et al., 2014). The IPAQ is one widely used PA questionnaire with two versions: a 39-question long form

and a 9-question short form (Hallal et al., 2012). While the IPAQ was originally validated in 12 countries and exhibits high test-retest reliability, it tends to overestimate PA levels and is weakly correlated with objective measures (Craig et al., 2003). The IPAQ has since been translated and tested in over 100 countries (Hallal *et al.*, 2012). Using a simple equation, it is possible to convert responses to the IPAQ into an estimate of the number of active METs undertaken by an individual over a normal week period, providing a single figure describing total physical activity undertaken. The limitations of the IPAQ include structure and framing of questions, low sensitivity compared to objective measures, and the requirement for PA to occur in bouts of 10 minutes or more (Rutten et al., 2002; Loyen et al., 2016).

However, objective physical activity measurement has become more common with the decline in monitoring cost and rise in subject convenience (Troiano *et al.*, 2014). The most commonly used objective measurement is accelerometery, with accelerometers used to measure PA in humans since the 1980s and 1990s (Plasqui et al., 2013). These devices have become smaller and lighter, causing minimal inconvenience to subjects. Accelerometers capture acceleration from a single point on the body (usually the waist or wrist) and convert it into electrical signals, called counts, which reflect the amount of force generated by the subject. Counts are summed over a specified time-period to provide a measure of activity intensity. Cut points can be specified to distinguish levels of activity (light, moderate, vigorous), however, there is no consensus on exactly what these cut-points should be (Lee & Shiroma, 2014).

There are several challenges in using accelerometery to measure PA, including limited ability to capture movements at the extremities, poor measurement of weight carrying, and limitations in measuring activities such as cycling (Lee & Shiroma, 2014). Pragmatic limitations also exist, such as subjects losing or breaking their accelerometers or forgetting to wear them. Use of accompanying surveys can mitigate these risks. Large datasets, such as the UK Biobank Study (Doherty *et al.*, 2017), are now available for cross-disciplinary research, however, the relationship between PA and health outcomes is generally based on subjective measurements and applying risk ratios from these studies to changes in objectively measured PA is unlikely to be reliable.

The gold standard for measuring energy expenditure is the Doubly Labelled Water (DLW) technique, which uses labelled isotopes to estimate average daily metabolic rate (Westerterp et al.,1986). However, it is an indirect measure of PA and does not distinguish between intensity, mode, or duration of physical activity. It is often used to validate other PA measurement methods (Maddison et al., 2007; Plasqui & Westerterp, 2007).

To standardise physical activity measurement, physical activity is typically categorised in terms of three levels of intensity: Light, Moderate, and Vigorous. The definitions of these levels vary slightly between organisations but generally involve faster breathing, increased heart rate, and difficulty speaking. There is no agreed standard mapping of these categories to METs, which makes meta-analyses and systematic reviews difficult (Daskalopoulou et al., 2017). However, there is a tendency to use 3-4 METs

for moderate intensity and 6-7 METs for vigorous intensity (Ainsworth et al., 2000; Haskell et al., 2007; Katzmarzyk et al., 2017).

Physical Activity Levels

Pooled analysis of surveys from over 122 countries with over a million participants find that around a quarter (27.5% and 22% respectively) of the world's population is physically inactive (Guthold et al., 2018; Dumith et al. 2011b), regardless of whether defined as "*not doing at least 150 min of moderate-intensity, or 75 min of vigorous-intensity physical activity per week*" (Guthold et al., p.1077) or "*engaged in less than 20 min/day of vigorous-intensity physical activity on at least 3 days/week, or less than 30 min/day of moderate intensity physical activity on at least 5 days/week, or less than 600 MET-min"* (Dumith et al. 2011b, p25), as shown in Table 2. There is significant variation across countries and WHO regions, with physical inactivity being higher in more developed regions (e.g. Western Pacific & Europe) than in less developed regions (e.g. South-East Asia and Africa). Dumith et al. (2011) which found a positive correlation between the Human Development Index (HDI) and physical inactivity in 76 countries. Both studies found that females were more likely to be inactive than males, an effect consistent in most countries (Guthold et al., 2018; Dumith et al. 2011b).

Table 2 Prevalence of insufficient physical activity among adults aged 18+ from the World Health Organisation Global Data Observatory, 2016, with 95% CI in brackets.

WHO region	Overall	Male	Female
Global	27.5 [25.0-32.2]	23.4 [21.1-30.7]	31.7 [28.6-39.0]
Africa	22.1 [19.9-24.0]	18.4 [15.8-20.9]	25.6 [22.8-28.2]
Americas	39.3 [37.4-40.9]	33.1 [30.8-34.0]	45.2 [42.9-48.6]
South-East Asia	30.5 [21.6-46.8]	22.9 [15.1-49.8]	38.3 [27.0-64.0]
Europe	29.4 [27.9-32.1]	26.2 [23.9-29.5]	32.4 [30.5-37.0]
Eastern Mediterranean	34.9 [32.1-39.2]	26.9 [25.4-30.6]	43.5 [41.4-46.6]
Western Pacific	18.6 [16.5-23.5]	18.8 [16.3-25.1]	18.5 [15.5-27.3]

Source: WHO GDO https://apps.who.int/gho/data/view.main.2482?lang=en

The IPAQ questionnaire is prone to overestimating physical activity (PA), and therefore the figures above are likely to underestimate physical inactivity (Craig *et al.* 2003). To address this, Loyen et al. (2016) analysed objective PA data from 5 European studies of 4 countries, finding that 72% of respondents did not meet WHO PA recommendations (not so different from the estimates above). England had the lowest level of PA among the 4 countries studied and the highest sedentary time. The

2008 Health Survey for England showed only 39% of men and 29% of women met the CMO's PA guidelines, and 6% of men and 4% of women met them based on objective measurements (Craig *et al.*, 2009). Children were found to be less likely to meet the higher levels of PA guidelines for children, with only 7% of boys meeting PA guidelines, and 0% of girls in the sample meeting guidelines of 60 minutes of PA per day in segments of 10-minutes or more.

Analysis of the trends of PA have been limited by the lack of comparable data (Hallal et al., 2012). Nevertheless, several findings have emerged from analysis of large datasets. Firstly, occupational physical activity has fallen substantially since the late 20th Century (Stamatakis *et al.*, 2007; Church et al., 2011), and secondly leisure time physical activity has increased since 2004, especially in groups with the lowest PA levels (Gu *et al.*, 2016; Craig et al., 2004). In a study of Swedish and Estonian children, Ortega et al. found that Swedish cohorts were less active at age 15 in 2004 than the first Swedish cohort were at age 15 in 1998. Finally, objective measures of physical fitness have generally been found to have fallen in most studies (Santtila *et al.* 2006; Peters et al. 2014; Eberhardt et al. 2020).

Determinants and Correlates of Physical Activity

The sub-field of public health aimed at understanding the determinants of risk factors such as smoking, alcohol consumption, and physical inactivity is relatively well established. However, physical activity behaviour differs from most other risk factors such as smoking and alcohol consumption since it requires an initiation of an action rather than ceasing it, for example stopping smoking or excessive alcohol consumption (Rhodes & Nigg, 2011). In the case of PA, the path of least resistance or inertia is the absence of the desired behaviour; it is not a variation on a necessary behaviour for survival (e.g. *healthy* eating); it requires a significant time commitment (vs tooth brushing, flossing, and sun-protective behaviour); and it is not a temporary one-time decision (vs cancer screening and radon testing); and it must be performed above the metabolic equivalent of rest sometimes with associated mild discomfort. However, it is similar to other risk factors in that there exist complex interplays between different determinants and the possibility of reverse causality and feedback loops at individual and population levels (Spence & Lee, 2003). The effects of some components are also temporal, making it difficult to determine cause and effect. Despite these difficulties, conceptual models have been developed to explain and predict PA levels for individuals and populations.

This section provides a review of sociological and ecological models of the correlates and determinants of physical activity. The terms 'correlates' and 'determinants' are used intentionally and have separate definitions. Correlates of physical activity are variables which have been shown to vary with changes in PA, generally in cross-sectional studies, while determinants of physical activity are variables which have been shown to have a causal relationship, generally in cohort or longitudinal studies. A weakness of the early physical activity literature has been the failure to distinguish between the two (Bauman *et al.*,

2002). Bauman *et al.* recommend that “the term “determinant” be used with greater precision and not be used to describe correlates of physical activity” (p.6).

There have been several systematic reviews undertaken over the past 20 years. Salis & Owen published the seminal comprehensive review in 2000, followed by Trost *et al.* who updated this in 2002. A more recent review, that of Bauman *et al.* (2012) builds upon the two previous works, using an adapted ecological model of the determinants of PA to structure a systematic review of the literature from 2000. This review paper was used to inform this review of the literature.

Gender and age are two significant predictors of physical activity levels, according to Bauman et al. (2012). Males are generally more active than females (Salis & Owen, 2000; Trost et al., 2002; Craig et al., 2009), with cultural differences affecting activity levels (Hovsepian et al., 2016). Physical activity levels typically fall with age in adulthood (Bauman et al., 2012; Craig et al., 2009; Grzywacz & Marks 2001). Dumith et al. (2011b) reviewed 26 cohort studies from developed countries and found a 5.9% annual decrease in PA levels from childhood to adolescence, with no significant gender difference in the rate of decrease. It is less clear whether there are differences in physical activity between ethnic groups, since investigation of these effects are generally confounded by the level of economic development in a country, socioeconomic inequality, and cultural practices.

Self-efficacy (SE) was reported as a correlate or determinant of PA levels in most of the systematic reviews included in Bauman *et al.* (2012). In the context of PA, an individual with high self-efficacy may be optimistic about their ability to achieve their target of 10,000 steps a day. Many systematic reviews of PA have shown self-efficacy to be a correlate of PA behaviour (Trost *et al.*, 2002). One study, that of Ishii *et al.* (2010) used structural equation modelling of the Japanese version of the IPAQ to create a model of the influence of non-biological factors on moderate PA. They show that non-biological factors influence PA through self-efficacy, such that self-efficacy is the mediator of the effect of every other variable on moderate physical activity. The role of self-efficacy as a mediator of other determinants of PA has been validated in numerous studies (Lewis *et al.*, 2002), and suggests that investments in other factors, for example the environment, may be poor value for money if self-efficacy is not also addressed.

People with higher education and income are more likely to meet PA guidelines (Loyen et al., 2015; Farrell et al., 2014; Gu et al., 2016). Education predicts physical activity levels independently of income, according to Cerin & Leslie (2008). Grzywacz & Marks (2001) found that less educated men tend to have a faster decline in activity levels over time. Higher socioeconomic status positively correlates with physical activity levels in developed countries (Trost et al., 2002). Although lower socioeconomic individuals face unfavourable PA environments due to less amenities, higher crime and lower walkability, the authors find that self-efficacy and social support account for most of the PA differences between the groups. There is some indication that this is in part hereditary, with adolescents whose parents reported being physically active throughout life being 6 times more likely to be physically active currently compared to those whose parents were not physically active (Christofaro *et al.*, 2018).

Farrell et al. (2014) analysed the correlates of PA in a sample of over a million respondents in the UK. The authors showed that PA increases monotonically with income and education, and the socioeconomic gradient of physical activity is not fully explained by differences in environment or access to resources. In the United States Gu et al. (2016) found that professional and technical staff are most likely to meet leisure time PA guidelines, while those in primary industries are least likely, using data from the National Health Interview Survey (NHIS).

In the Bauman *et al.* review several environmental factors are identified as being potentially important, although for all of them the evidence is mixed. Findings from their review suggest that objectively measured variables such as walkability, traffic speed and volume, proximity to green space, and access to recreational facilities are the most important environmental variables for children. A finding validated by Sallis *et al.* (2016) who assessed objectively measured (using GIS) environmental correlates of objectively measured physical activity in their study of 14 cities. This suggests an urban/rural divide with those in more rural areas with less walkability more likely to be inactive.

Giles-Corti & Donovan (2002) compares the relative importance of distinct types of determinants (individual, social environmental and physical environmental) of physical activity. When assessing relative importance of individual vs environmental factors, it appeared that physical environmental factors were the least important.

Physical Activity over the life-course.

Physical activity decreases and sedentary behaviour increases with age, as shown by cross-sectional and longitudinal studies (Bauman et al., 2012). Critical points in the life-course, related to life events, may exacerbate these declines (Varma et al., 2017). Examples include biological and sociocultural changes in adolescence (Craggs et al., 2011), pressures of work and childcare in early adulthood (Allender et al., 2008), mid-life caring responsibilities and then later-life health limitations (Sun et al., 2013). Minimising reductions at these critical points is important for public health.

Physical activity tends to decrease particularly rapidly in childhood and adolescence in both boys and girls in developed countries, regardless of whether measured objectively or subjectively (Ortega et al., 2013; Dumith et al. 2011a; Craggs et al. 2011). Dumith et al. (2011a) conducted a systematic review of 26 studies on PA change from childhood to adolescence in developed countries, finding a mean annual decrease of 5.9%, with only one study reporting an increase in PA level for boys. The authors observed no significant difference in results between genders but noticed higher declines in PA for girls in recent studies, as confirmed by Craggs et al. (2011). The fall in PA in adolescence is particularly important because many behaviours become habits in adolescence and these habits are more stable in adulthood (Lounassalo et al., 2019).

National Physical Activity Strategies & Targets

International organisations such as the World Health Organization (WHO) provide global guidance on physical activity strategy and recommendations. The WHO's Global Strategy on Diet, Physical Activity and Health was adopted in 2004, providing a framework for developing a "good" national strategy for physical activity (PA). The framework includes areas such as national leadership, supportive environments, policies, programs and monitoring. The *Global Recommendations on Physical Activity for Health* was included in the 2008-2013 Global Strategy for the Prevention and Control of Non-communicable Diseases (WHO, 2013). They were created to provide a simple set of guidelines which could be marketed at a global level or used to inform national guidelines, as has occurred in the UK. These recommendations are currently being updated, but the latest recommendations by age category can be found in Bull et al. 2020.

Devolved authorities within the UK have made numerous efforts to increase physical activity (PA) through various public health strategies. In 2014, Public Health England (PHE) published "*Everybody Active, Every Day*" which reported that the UK population was 20% less active than in 1961. The report suggested ways to improve activity levels by changing the culture, environment, and scaling up effective interventions. In response, the UK government committed to funding inactivity prevention through Sport England's "*Towards an Active Nation*" strategy, dedicating 25% of resources to tackle inactivity and setting targets for inactive people to benefit. However, a 2017 report by PHE found no objective improvement in PA levels. Scotland has set a more objective target of 50% of adults and 80% of children meeting minimum PA levels by 2022. The Scottish government has also published "*A More Active Scotland*" to build on the legacy of the 2014 Commonwealth Games and improve PA through changes in physical environments and investments in sport and recreation.

The UK's physical activity guidelines, as well as giving a definition for inactivity, are themselves an intervention - aiming to advise the population on healthy behaviours. First introduced in 1996, they recommended a minimum of 30 minutes of moderate to vigorous PA on at least five days a week. The latest guidelines from 2019 recommend that children aged 5-18 should "*engage in moderate-to-vigorous intensity physical activity for an average of at least 60 minutes per day across the week*" and that adults aged 19-64 should "*accumulate at least 150 minutes (2 1/2 hours) of moderate intensity activity (such as brisk walking or cycling); or 75 minutes of vigorous intensity activity (such as running)*" or combinations thereof (Department of Health and Social Care, 2019).

Despite social marketing campaigns like Change 4 Life, public knowledge of physical activity guidelines remains limited. Surveys in 2007 and 2013 showed only 11-18% of respondents had accurate knowledge of the guidelines (Knox et al., 2013). There is a social gradient in awareness, with older men and those with lower education being less likely to know the guidelines. The authors suggest that inconsistent messages from various campaigns may have led to confusion and misinformation.

Health Economic Evaluation

A health economic evaluation refers to a study that estimates the incremental costs and health outcomes of one or more interventions relative to another, often the current strategy (Drummond et al. 2015; Briggs et al. 2006). There are several different types of economic evaluations, the most used being cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA). A CEA compares the incremental costs and health effects of an intervention in a common unit, for example: hospital admissions avoided, symptom free days, or cases of a disease avoided. A CUA is a subset of CEA, using a generic outcome measure, such as quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) which combine duration and health-related quality of life (utility) into a single index. While this allows for direct comparison across outcomes, the indices are typically more difficult to estimate than the simple common unit. In both cases, an intervention's incremental cost per outcome metric, the additional cost per unit of the outcome relative to the next best intervention, is compared to the willingness to pay (WTP) threshold of the decision-maker. Where an intervention provides incremental health outcomes at a cost below an acceptable threshold, which in theory is determined by the opportunity cost in terms of activities displaced by making additional funding available for the intervention, the intervention will be determined to be 'cost-effective'.

CBA is an extension of CEA and CUA, which converts both incremental costs and health outcomes into monetary values, enabling a direct comparison between the incremental costs and benefits of health-related interventions with other interventions (e.g., in education or transport). However, this creates an additional challenge of explicitly determining an appropriate monetary value for health outcomes, including the duration and quality of human life, at the modelling stage, rather than at the decision stage as with threshold analysis used for CEA and CUA.

Since data is often limited and many costs and outcomes accrue in the future, health economic evaluations often require the development of a decision model. A health economic decision model is a computational abstraction of the real world designed to provide an estimate of the outcomes achieved by different courses of action. This computational model, typically created using computer software, is generally a simplification of a conceptual model, a qualitative summary of human understanding of real-world systems. Its primary purpose is typically to predict the outcome from different courses of action and/or to incorporate preference weights or valuations for different trade-offs, to inform decision-making processes relating to health.

The extent to which a model attempts to capture the nuances of the real world can vary significantly. A simple computational model might involve a trade-off between two courses of action, each of which have a probability of success and failure and a corresponding one-time payoff, for which the expected values could be calculated with simple mental arithmetic. A sophisticated computational model might capture the dynamic transmission of an infectious disease, incorporating factors such as contact patterns, spatial distribution, and the effects of vaccination and mitigation measures and may

summarise the costs and benefits of different interventions using elicited societal preferences for different outcomes.

However, using models to inform policy has many potential limitations. Firstly, a model is limited by the extent to which its parameter and data inputs accurately reflect the real world. Many inputs to a model will be uncertain, which when combined can result in considerable uncertainty in its predictions. Sensitivity analysis can help identify the potential impact of uncertainty in parameters for which the modeller is conscious of this uncertainty (*known unknowns*), but it cannot eliminate it. Moreover, the implications of overconfidence in the presence of *unknown unknowns* can be substantial, for example where data used to inform the model indicates relative certainty because of biases in the wider literature.

By their nature, models simplify the real world, meaning that their structure may not fully capture the intricacies of the system that they are trying to better understand, for example a healthcare system or population. This can limit their ability to provide reliable estimates of the effect of interventions on health outcomes and costs. This is often the case in systems which involve complex interactions, nonlinear relationships, and feedback loops (Breeze et al., 2023). These systems are characterized by a high degree of uncertainty, dynamic changes, and interconnected factors that can be challenging to model accurately. Consequently, the limitations in the model's structure may lead to incomplete or even misleading insights and make it difficult to understand the external validity of the model, resulting in limited external validity in other contexts outside of the primary decision, for example when applied in different countries of for different groups within society.

Finally, there are many cases that computational models cannot, in themselves, provide decision-makers with a dominant decision, one in which one course of action results in the most positive outcome for every consideration. As a result, decision-makers still need to weigh the trade-offs between different outcomes or, provide a quantitative value for the trade-off to be incorporated explicitly in the model. This is often very difficult to elicit explicitly from decision-makers, as was particularly marked in decisions by governments worldwide on how best to respond to the COVID-19 pandemic.

Nevertheless, models can provide valuable insights to inform policy decisions, so long as all assumptions, structural motivations, and outcome uncertainties are clearly communicated, and results are interpreted with caution. At very least, they can provide a transparent means by which decision makers can clearly articulate the logic behind specific courses of action.

The application of health economic evaluations (and models) to public health, however, is still relatively new and many challenges exist, including a less formal modelling framework compared to the evaluation of pharmaceutical products and the absence of randomised controlled trials (Squires et al. 2016b). As a result, health economic modellers working on economic evaluations of physical activity interventions have considerable flexibility in model design and must make a series of structural assumptions that carefully balance the external validity of the model with its ease of use and ability to inform decision-making (usability). Understanding the impact of these structural assumptions is crucial

for promoting good practice in health economic evaluations of physical activity models and for effective resource allocation in the long term.

What is external validity?

A model's validity refers to the extent to which it accurately reflects the real world, or phenomenon it is intended to simulate . A 'valid' model should produce reliable predictions about the real-world system it is trying to simplify. To achieve this, it will generally have a structure, set of assumptions and parameter inputs based upon the best available understanding of the world (Drummond et al., 2015; Briggs et al. 2006).

There are different aspects of model validity, but the most often considered include face validity, internal validity, and external validity. Face validity typically refers to a subjective assessment of whether a model accurately captures the important components of the system it is trying to simplify. Internal validity describes the extent to which a model '*makes sense*' in and of itself, matching the understanding and intentions of its creator. For example, a model of disease progression would not be internally valid if it did not closely match that observed during the clinical trial from which it was informed. External validity refers to the model's generalizability and applicability to different populations or settings beyond the specific conditions for which it was initially developed. For example, a model may accurately predict health outcomes during the clinical trial for those included in the trial but not in the years following a trial or for a different group of people. It is external validity that is generally of greatest concern for decision-makers, and that is the focus of much of this thesis.

Assessing validity is important to ensure that the model produces results that can be relied upon to help inform decisions (McCabe & Dixon, 2000). However, since health economic models are typically necessary due to limitations in observed data, it can be challenging to assess their external validity quantitatively. However, it is possible to assess the credibility of both the internal mechanics of models against the research community's conceptual understanding of relationships observed in the real world (face validity). It is possible to review the model for error in calculations and internal logic (model code review). It is also possible to compare the results of the model to the results of other models trying to solve the same decision problem and understanding reasons for differences (double coding or benchmarking). In some cases, for example in retrospective studies, or for parts of the model for which there is good data, the performance of the model can be assessed quantitatively against observed outcomes (cross-validation). The thesis discusses the external validity of existing health economic evaluations of physical activity interventions in Chapter 2, while Chapters 3-6 focus on methods for improving external validity.

Establishing model validity is crucial for building confidence in the model's results. As a result, the impact of a model on decision-making is highly dependent on its perceived external validity. However, another characteristic of models which effect their impact on decision-making is their *usability*.

What is usability?

In the context of this thesis, usability refers to the extent to which the model improves the understanding of the decision problem to effectively inform a decision. Usability is important to ensure that decision makers can understand the implications of their resource allocation decisions. A model is more usable if it is accessible, transparent, up to date, flexible, and able to accommodate different scenarios and assumptions, well documented, simple to understand, and provides the information required by the decision maker. Chapters 7 and 8 focus on ways to improve the usability of health economic models, including strategies for making models more accessible, transparent, and adaptable.

The following chapter reviews the methods used in health economic evaluations, and subsequent chapters of this thesis test the effects of some of these assumptions on key outcomes, while considering the trade-off between external validity and the usability of the model.

Chapter 2. Scoping review of health economic evaluations of physical activity interventions

Introduction

The aim of this scoping review is to assess the methods typically used in economic evaluations of physical activity interventions and to identify limitations and key assumptions to inform the development of a new model with greater external validity. In this context the external validity of a model is the extent to which the model captures our best understanding of the mechanics of the decision problem in the real world.

The review was originally undertaken in August 2017, but an additional section at the end of the chapter provides an update for the period from August 2017 to January 2023.

This review is divided into two parts. The first part focuses on economic evaluations that were conducted to inform the guidance issued by the National Institute for Health and Care Excellence (NICE) in England. NICE guidance is widely used by the National Health Service, Public Health, and local authorities to make decisions, and it is also referenced by other countries. Economic evidence plays a crucial role in the NICE decision-making process (Dakin et al., 2015; Tappenden et al., 2007), therefore, the methodology employed in these economic evaluations is expected to significantly influence NICE guidance and subsequent decision-making.

The second part of this review expands to examine economic evaluations published in peer-reviewed journals to determine if there are any advancements in the literature that have not yet been incorporated into NICE submissions. Throughout the review, there is a specific emphasis on evaluations that primarily focus on children and adolescents, to determine if there are any challenges unique to evaluating the economic impact of physical activity interventions for this population.

This review focuses on published economic evaluations and should not be considered an exhaustive critique of all evaluations that have been conducted. Many economic evaluations remain unpublished, particularly those carried out by consultancy firms that are not affiliated with a university.

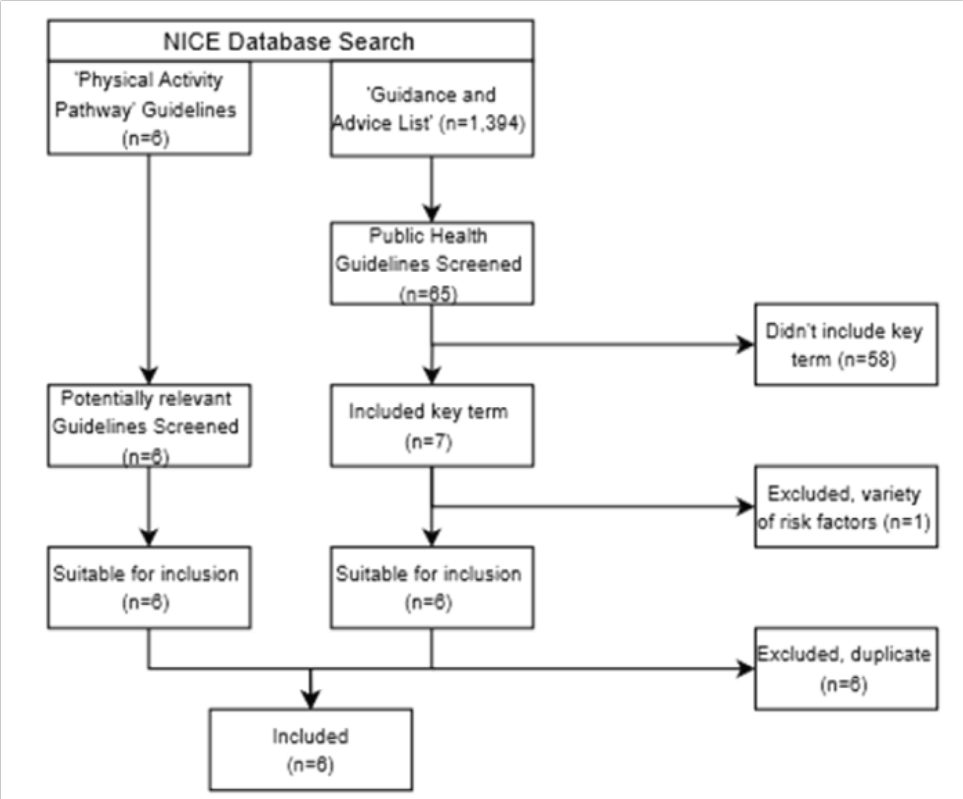
NICE Physical Activity Models

Methods

The focus of this section of the review is restricted to NICE public health physical activity models published in the last 10 years (2007 to 2017). To identify relevant NICE guidance, two searches were

conducted. The first search was within NICE's online *Physical Activity Pathway directory (https://www.nice.org.uk/guidance/lifestyle-and-wellbeing/physical-activity)*, and the second search was in the NICE full *Guidance and Advice list* (https://www.nice.org.uk/guidance/published) using the key terms "*physical activity*", "*exercise*", and "*active travel*". For each physical activity guidance identified, the evidence section was examined, and the related economic report was retrieved. During the review of each economic report, any references to other NICE guidance were followed up using a snowball approach to ensure that no guidance was missed. Any relevant reviews discovered during the snowball approach were added to the review of the published literature (Part 2). A figure of the search strategy can be seen below.

Figure 1 NICE Physical Activity Model Search Strategy Flow Diagram



For each economic report a data extraction form (see supplementary material for form row titles) was used to ensure consistency in the review. The data extraction form was based on a combination of the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist (Husereau et al., 2013), a review of public health models (Squires et al., 2016a), and several items which emerged from the *economic reports*. After all *economic reports* were reviewed, the reports were reviewed again in reverse order to ensure no bias emerged from ordering.

Results

There were 65 Public Health guidelines published by NICE as of 11/10/2017. Of those guidelines, 6 were included within NICE's '*Physical Activity Pathway*' and 7 included the words '*Physical Activity*', '*Exercise*' or '*Active Travel*'. Six of the guidelines were duplicates, the additional guideline in the second search compared many interventions targeted at improving mental health of over 65s, many of which were not related to physical activity, and as such was excluded from the review (Windle et al. 2007). The remaining 6 guidelines relate directly to physical activity (Table 3 below), covering different areas and populations.

Table 3 NICE Public Health Guidance for Physical Activity Interventions

ID #	Name	Author
PH13	Physical Activity in the Workplace	Bending et al. (2008)
PH17	Physical Activity for Children and Young People	Fordham & Barton (2009)
PH44	Physical Activity: Brief Advice for Adults in Primary Care	Anokye et al. (2012)
PH41	Physical Activity: Walking and Cycling	Brennan et al. (2012)
PH54	Physical Activity: Exercise Referral Schemes	Campbell et al. (2013)
PH8	Physical Activity and the Environment	Love-Koh & Taylor (2017)

Model topics included: individual level interventions (PH44, PH54), community interventions (PH17, PH41, PH8) and workplace-based interventions (PH13). At the time of the review, there was no NICE model that evaluated school-based interventions. The published guidelines were developed by a committee at NICE, which analysed the results of an economic model for each decision problem. This section describes each model, and the discussion section that follows analyses the key similarities and differences among them.

Bending et al. (2008)

In their paper, Bending et al. performed an economic analysis of four workplace interventions aimed at promoting physical activity among sedentary adults. These interventions included: light-touch counselling, more extensive counselling, a walking programme and a fitness programme. The authors calculated the incremental cost-effectiveness ratios (ICERs) by comparing the cost and benefits of each intervention to a "do-nothing" scenario with zero costs and benefits. No incremental analysis was performed to compare the interventions.

The authors estimated the ICERs using parameters from four separate studies by Østerås et al. (2006), Aittasalo et al. (2004), Chyou et al. (2006), and Purath et al. (2004). They used a cohort model based on data from the Health Survey for England 2004 (HSE, 2004) to simulate the risk reductions of three conditions: CHD, stroke, and type 2 diabetes, based on being in one of five physical activity categories.

The model does not consider changing levels of physical activity over the life course or the decay of intervention effects, instead assuming that "*physical activity levels are maintained over a period sufficient to ensure that the health benefits associated with that level of activity are attained*" (p.6). There is no consideration of sub-group effects or of impact on inequalities in health outcomes.

The authors also included the impact of physical activity on absenteeism in their analysis, informed by a single study evaluating a different fitness program over a decade earlier, leading to highly uncertain results. The ICERs for the workplace interventions ranged from £495.50 to £1,234.11 per Quality Adjusted Life Years (QALY) gained.

Fordham & Barton (2008)

Fordham and Barton evaluated the cost-effectiveness of four interventions aimed at enhancing physical activity among young people: walking buses, free swimming, dance classes, and community sports. The cost-effectiveness of each intervention was assessed using the incremental cost-effectiveness ratio (ICER), which compared the costs and benefits of each intervention to a "do-nothing" scenario with no costs or benefits. No incremental analysis was undertaken.

The authors relied on the evidence of effectiveness from a NICE review for the walking bus intervention (NICE, 2007). For the other three interventions, the absence of sufficient effectiveness evidence necessitated relying on several assumptions. As a result, the estimates of increased physical activity are based on assumptions about displacement of other physical activity and drop-out rates derived from studies performed in various settings, which may not be relevant to the UK (Kingham et al. 2005). The costs for each intervention were estimated using policy documents.

The evaluation employs a simplified decision tree type model, where the quality-adjusted life years (QALYs) improvements for each participant in an intervention are calculated by multiplying the additional minutes of physical activity by an estimate of "*Gain of PA per minute*" (approximately 0.0001 QALYs) from Beale et al. (2012). This eliminates the requirement for a baseline population for comparison. Consequently, although physical activity is treated as a continuous variable in the model, it does not account for the non-linearity of physical activity benefits, effects on specific health conditions, subgroup analysis, or considerations of inequalities. The model does not include any non-health benefits.

The authors acknowledge the limitations of the data and the model's reliance on numerous assumptions. They state that "due to the limitations of the available evidence, our analysis did not seek to estimate the long-term cost-effectiveness of these interventions" (p.28) and that "there is a large amount of uncertainty associated with these results as, due to the limitations of the evidence, it was necessary to make a number of unverified assumptions within the analyses" (p.2).

The ICER for the interventions varied from £4,007.63 for the walking bus intervention to £71,456 for the community sports intervention, as compared to the "*do nothing*" scenario. However, the model failed to

furnish decision-makers with confidence. In issuing the final guidance for PH17, NICE noted that "There is virtually no evidence on the cost-effectiveness of interventions to increase children and young people's physical activity levels" (NICE, 2009).

Anokye, Lord & Fox-Rushby (2012)

Anokye, Lord & Fox-Rushby model the cost-effectiveness of brief advice to inactive adults in a primary care setting relative to usual care with no active intervention, from the perspective of the National Health Service (NHS). They use parameters from a meta-analysis of the effectiveness of brief advice (Pavey *et al.* 2011), relative risks for coronary heart disease (CHD), stroke, and Type 2 Diabetes (Hu et al. 2007, 2005, and 2003, respectively), and costs from Boehler et al (2011). The study distinguishes itself from others by incorporating short-term mental health benefits of physical activity into the model.

The structure of the model is based on Anokye et al. (2011). A cohort of 100,000 inactive individuals, not meeting the Chief Medical Officer's guidelines, aged 33 either become active or remain inactive after one year. For the remaining 48 years until their death, they maintain the risks associated with this dichotomous physical activity status. The authors argue that modelling changes in physical activity over the lifetime is not necessary because epidemiological evidence is based on cohort studies that span long periods of time, during which physical activity status may be transient. The authors do not consider that individuals who become active for one year are more likely to revert to an inactive status than those who have been consistently active. The model does not take into account non-health benefits or equity considerations.

The authors of the study make several recommendations for future research including a particular focus on "*the nature of mental health gains (size and duration) from physical activity participation, and when and how they can be measured*" (p44). Other recommendations include the need for better quality evidence of the effectiveness of physical activity interventions for subgroups and in particular on the decline of effectiveness over time.

The base case estimate for the ICER was £1,730. However, the ICER was highly sensitive to the inclusion of short-term mental health gains. Excluding these gains increased the ICER to £27,000.

Brennan et al. (2012)

Brennan et al. evaluated the cost-effectiveness, measured in terms of pounds per quality-adjusted life year (QALY), of four interventions aimed at promoting walking and cycling in adults for transportation or recreation purposes. The interventions included two programs that aimed to increase the use of active transport, a pedometer-based program, and community-led walks (Sloman, 2009; Cobiac, 2009; Fordham & Barton, 2009). The primary focus was on "health economic" outcomes (Brennan et al., 2012, p.142), supplemented by additional analysis of the environmental and economic benefits of the interventions.

The model is an individual simulation model, CD2 in Brennan's taxonomy (see Brennan et al., 2006), which treats physical activity as a continuous variable, combining estimates of increased minutes of PA in the population with a non-linear dose-response function between minutes of activity and all-cause mortality derived from ordinal estimates from Anderson (2000). The baseline population of the model was informed by the population aged over 16 in the Health Survey for England (HSE, 2008), and was used to estimate risk reductions from small increases in physical activity. The model is also the first NICE model to incorporate the effect of increases in one type of PA (e.g. cycling) on another (e.g. walking) which allows for consideration of displacement effects.

The authors acknowledge several limitations of the model, including the assumption that risks change immediately (next cycle) following changes in PA behaviours. Also, while decay of effect is varied in sensitivity analysis (and found to be very important) over the entire range, from 0% to 100%, there is no sensitivity analysis of the proportion of individuals developing long-term habits independent of this decay. No subgroup analysis or distributional analysis was reported. A further limitation is that no diseases are included, only all-cause mortality is modelled, a similar limitation to other transport models such as the HEAT (Kahlmeier et al. 2017).

The estimated ICERs relative to doing nothing ranged from £300 per QALY gained for Travel Smart to £5,000 per QALY gained for the Cycling Demonstration Towns intervention.

Campbell et al. (2013)

Campbell et al. conducted an economic evaluation of exercise referral programs aimed at inactive but healthy individuals in primary care, compared to no intervention. The evaluation built on the NICE model from Anokye et al. (2012) discussed above which considered the effects of exercise referral schemes on CHD, Stroke, and Type 2 Diabetes. The evaluation incorporates estimates of the efficacy of exercise referral programs estimated from a review by the same authors and incorporates updated healthcare costs (Campbell et al., 2013).

The model is structured as Anokye et al. (2012). It follows a cohort of 100,000 individuals who are inactive at age 50 and can become active after one year if the intervention is effective. Once they adopt a physical activity status, they maintain the corresponding risks for the rest of their life. The model includes subgroup analyses for individuals with hypertension, obesity, or depression, but does not account for health outcome inequalities.

The authors highlight several limitations of the model, including the assumption that the efficacy of physical activity interventions is uniform across the population, excluding those with depression, which is unlikely to be accurate. Additionally, the model treats conditions as mutually exclusive and limited to three, likely resulting in an underestimation of overall QALY gains. Despite its limitations, the model makes an important contribution to the literature by considering comorbidities and different risk ratios

for specific subgroups, such as those with depression and obesity. This approach has the potential to be expanded to a larger number of comorbidities, enhancing the validity of future models.

However, an important contribution of the model to the development of the literature is the consideration of comorbidities and differences in risk ratios for specific subgroups within the model, most notably by depression and obesity. This method could be further developed for a substantial number of comorbidities to improve the validity of future models.

The ICER for exercise referral schemes, for inactive but otherwise healthy individuals, versus 'do nothing' is estimated to be £76,276.

Love-Koh & Taylor (2017)

Love-Koh and Taylor undertook an economic evaluation of eight interventions aimed at improving the physical activity environment for the general population - evaluating each against a do-nothing scenario. They calculate baseline PA distributions for each age-group measured in units of metabolic-equivalent time (MET) from the Health Survey for England (HSE, 2014). The model employed non-linear dose-response functions for five conditions (CHD, Stroke, Diabetes T2, Colon Cancer, and Breast Cancer) from Kyu et al. (2016). However, the rationale for choosing these specific conditions is not fully explained. The authors acknowledge the limited scope as a limitation and included mortality risk reduction data from Anderson et al. (2000) to provide a more comprehensive measure of health.

The model structure was innovative and solves many of the limitations of the previous NICE models. It is closest to Brennan *et al.* (2012) in approach. After establishing a baseline distribution of PA, measured in METs, derived from HSE 2014, the increase in PA estimated from the eight studies were applied either equally across the PA distribution or unequally using a technique developed by Minton *et al.* (2013). While gender and age are included to improve the external validity of the model there is no separate subgroup analysis by individual characteristics other than whether individuals had limited mobility. Similarly, although the effect of changes in the distribution of PA increases across the PA spectrum on overall cost-effectiveness are modelled, there is no consideration of outcome inequity.

The authors acknowledge multiple limitations of the model. The use of effectiveness evidence only allowed for estimation of the average impact of the interventions on PA, precluding subgroup analysis for groups other than those with limited mobility. The biggest limitation, however, is the challenge in estimating the decay in the effect of the interventions. The base case assumes no decay, and a sensitivity run was conducted where a portion of the population experienced a fixed decrease in the intervention effect in the first period with no further decay. Finally, the sensitivity analysis is run on a deterministic model rather than on the probabilistic model.

The ICERs, compared to a do-nothing scenario, are shown to range between £1,397 for the Active Living Programme to £215,989 for park renovations.

Discussion

Several recurring themes emerged from the review of the NICE models including: (1) heterogeneity in measurement/categorisation of physical activity levels, (2) lack of focus on population heterogeneity, (3) limited scope of models to health and narrow outcomes within health, (4) limited consideration of health inequalities, and (5) limited or no modelling of intervention effect decay or tracking of physical activity over the life course. These themes generally stem from limitations in the quality of effectiveness studies and data, and limitations of existing modelling structures. However, recent advancements in the evidence base and in computing power may result in models based on more granular data with more externally valid modelling structures and a better incorporation of parametric and structural uncertainty analysis. A discussion of the quality of effectiveness evidence, the specific limitations of the NICE studies, and potential solutions is included in the following sections, with recommendations shown in bold text.

Limitations in the quality of effectiveness evidence, i.e. studies which measure the effect of an intervention on activity levels, is a recurring theme in the six physical activity models submitted to NICE. The effectiveness evidence for PA interventions is often based on small scale pilots which are evaluated with limited resources. However, even in larger scale studies in the published literature there are multiple limitations, including a lack of consistency in the measurement of physical activity, a lack of reporting differing effectiveness by subgroup, and short follow-up periods. This was highlighted explicitly in many of the modelling reports for example by Brennan et al. (2012) which highlights the need for *"better quality evidence on effectiveness of physical activity*", and Love-Koh & Taylor (2017) which had no alternative to using four effectiveness studies which had been given a minus score by the team at NICE.

Another common problem is that the definitions of physical activity (PA) used in studies evaluating the effects of an intervention effectiveness often differ from the definitions used in epidemiological studies. This mismatch makes it challenging for health economists to link intervention effectiveness to disease and mortality risk reduction. The use of METs obtained from questionnaires like the IPAQ Short Form may become the common yardstick that solves this problem. Shifts towards objective measurement of PA also provides a chance for researchers to standardise on a single measure or more easily convert between different measures.

The available evidence often lacks consideration of confounding variables or subgroup analysis, which would provide insights into the cost-effectiveness for different populations and the impact of physical activity (PA) interventions on inequalities. As noted by Bending et al. (2008), *"there is a lack of available evidence in the form of long-term natural history studies of PA interventions to understand how compliance may be influenced by factors such as age, lifestyle, profession, etc.*" (p.16). Campbell et al. (2013) aimed to analyse more subgroups, but "*subgroup-specific effectiveness data were only available for the depression subgroup*" (p.118). They also highlighted "*limitations in the clinical effectiveness evidence base*" (p.10).

Additionally, the limited follow-up periods in effectiveness studies result in the need to estimate long-term projections from brief periods, often no longer than 1 year. This leads to either significant uncertainty in the model results or to model assumptions that significantly impact the results but are based on limited evidence.

Effectiveness studies could be of better use to economic modellers if they: used standardised measurements of physical activity, increased sample size and included subgroups and equity effects, and increased follow up periods to provide more certainty in projecting long term effects. Modellers could provide more detail as to appropriate subgroups, length of studies, run sensitivities around intervention effect decays, and apply pressure on the community to standardise measurements.

Theme 1. Heterogeneity in measurement/categorisation of physical activity levels.

Table 4 summarises the form and units of measurement for the physical activity (PA) variable in the six NICE models. Three models (Bending et al., Anokye et al., and Campbell et al.) use a binary variable to describe PA, while the remaining three (Fordham & Barton, Brennan et al., Love-Koh & Taylor) treat PA as a continuous variable. Only the most recent model by Love-Koh & Taylor uses METs to measure PA, whereas earlier continuous models (Brennan et al. and Fordham & Barton) used a simpler measure of minutes of PA to calculate QALYs.

Table 4 Choice of units in six NICE commissioned cost-effectiveness models, 2008-2017

Model	Type of Model	Type of variable	Units
Fordham & Barton (2008)	Decision Tree	Continuous	Minutes of PA
Bending et al. (2008)	Cohort	Categorical	Active/Sedentary
Brennan et al. (2012)	Individual	Continuous	Minutes of PA
Anokye et al. (2012)	Cohort	Dichotomous	Active/Inactive
Campbell et al. (2013)	Cohort	Dichotomous	Active/Inactive
Love-Koh & Taylor (2017)	Cohort	Continuous	METs

Anokye *et al.* (2012) stated that "*in agreement with NICE, the binary outcome had to be used because of a lack of effectiveness data on people exercising below or above the threshold*" (p.14). Their model uses a binary physical activity variable whereby individuals are described as either inactive or active, where being active is defined by the authors as meeting or exceeding the government target of 150 minutes of at least moderate intensive physical activity or 75 minutes of vigorous activity. Bending *et al.* (2008) treat PA as a simple dichotomous variable, with individuals being described as either sedentary or active. All individuals enter the model as sedentary and become active with some probability P(a). Once categorised as 'active', individuals have an immediate reduction in their risk of CHD, Stoke and Diabetes compared to sedentary individuals. Campbell *et al.* (2013) treat physical activity as a

dichotomous variable, with individuals described as either active or inactive and are assigned risks for CHD, Stroke and Diabetes upon this basis. The authors argue that since large scale cohort studies have been undertaken linking activity level at baseline to long term risks, changes in PA levels over the life-course are already accounted for.

Fordham & Barton (2008) use the findings of Beale *et al.* (2012) linking PA to health benefit to estimate the QALY gain associated with an increase in physical activity. Beale *et al.* calculated that 6 hours of additional PA per year equates to a QALY gain of 0.0026692. Fordham & Barton make several assumptions, 1) that the distribution of this additional QALY gain across the year does not matter, 2) that the health benefit from additional PA is homogenous between groups, 3) that the health benefit is independent of baseline activity levels, 4) that intensity of PA does not matter 5) That the benefits from PA are homogenous over the life-course. Also, by linking PA directly to QALYs, ignoring specific conditions, the authors are unable to estimate cost savings associated with treating various diseases or describe the mechanisms via which their benefit estimate is derived.

Brennan *et al.* (2012) treat physical activity as a continuous variable, using a simple measure of minutes of PA per week derived from the HSE (2008). The authors estimated the effect of various interventions on the quantity of walking, cycling and driving given an individual's access to vehicles and their working status. This generated an increase in overall minutes of PA which was combined with a modelled continuous risk function from Anderson (2000), to generate a health benefit in life expectancy, deaths and QALYs. Love-Koh & Taylor (2017) is the first to treat PA as a continuous variable measured in METs. This is important because it ensures that intensity of PA is considered, consistent with the findings of the literature. This was made possible by a meta-analysis published by Kyu *et al.* in 2016 which estimated the relationship between METs/week and the relative risk of five different PA related diseases: Type 2 Diabetes, Stroke, CHD, Colon Cancer and Breast Cancer. This creates a far more detailed picture of the health benefits of physical activity for these diseases, although future models could improve upon this by including other diseases such as mental health conditions which are increasingly prevalent and have been linked to PA. The primary limitation with this method is that many effectiveness studies do not yet provide this granularity of data, and therefore increases in PA are provided as a simple average MET increase for the entire population, which means that the methodological improvement yields very little increase in the overall quality of the model.

Physical activity should ideally be modelled as a continuous variable in models, ideally in METs, to align with the epidemiological evidence indicating that health benefits of physical activity experience non-linear diminishing returns and that "some is good, more is better" (Department of Health and Social Care, 2019).

Theme 2. Lack of focus on population heterogeneity.

The six NICE models had diverse structures, resulting in varying base populations. Table 5 lists the model type, base population, and population charactersitics in each model. Four models (Bending et

al., Anokye et al., Campbell et al., Love-Koh & Taylor) were cohort models, one was a basic decision tree (Fordham & Barton), and one was an individual simulation model (Brennan et al.).

Table 5 Base Population used in the 6 NICE Physical Activity Models

Model	Type	Base Population	Variables	Subgroups	Source
Fordham & Barton (2008)	Decision Tree	Simply number in each intervention.	None.	None	N/A
Bending *et al.* (2008)	Cohort	1,000 sedentary individuals aged 40-65.	Age, prevalence of smoking, systolic blood pressure, cholesterol, BMI, waist circumference.	None	HSE (2004)
Brennan *et al.* (2012)	Individual	HSE (2008) population of 22,623 people.	HSE: Age, sex, work status, car access, minutes active. NTS: Trips, Cycling, Walking, driving (by purpose).	None	HSE (2008) NTS (2002-2008)
Anokye *et al.* (2012)	Cohort	100,000 inactive but healthy individuals aged 33.	Active / Inactive.	None	N/A
Campbell *et al.* (2013)	Cohort	100,000 inactive but healthy individuals aged 50.	Active / Inactive	Depression, Obesity, Hypertensive	N/A
Love-Koh & Taylor (2017)	Cohort	(# unspecified) based on HSE (2014) PA and age distribution.	Age, Gender, METs.	Limited Mobility	HSE (2014)

Cohort Models

The base populations in the cohort models vary. For example, Anokye et al. (2012) and Campbell et al. (2013) start with 100,000 inactive individuals aged 33, and 50 respectively whereas the Bending et al. and Love-Koh & Taylor models start with a population of mixed age. Some models include additional variables besides age and physical activity status to estimate risk more accurately. Bending et al. includes sex, smoking prevalence, blood pressure, cholesterol, Body Mass Index, waist circumference, and blood pressure medication usage, while Love-Koh & Taylor only includes gender. Anokye et al. and Campbell et al. do not include any other variables. Anokye et al. recommends that future research should develop a "*population model that accounts for a range of patient (and potential provider) characteristics and is able to consider, more directly, information from infrastructure-based interventions that influence access to services*" (p.44).

There was also variation in the subgroups analysed. The Anokye *et al.* and Bending *et al.* models do not include any analysis of subgroups. Campbell *et al.* obtained different risk ratios from the literature

for those with hypertension, obesity and depression, and distinguished between effectiveness for those with depression, and undertook subgroup analysis for the three sub-groups. Love-Koh & Taylor undertakes subgroup analysis for those with limited mobility.

The Fordham & Barton study was the simplest. The base population used is simply the number of children who undertook the intervention. There were no other variables incorporated other than the number of participants, since the benefit of the intervention was assumed to be homogenous for all individuals. There is also no consideration of subgroups or heterogeneity of effect in the population.

Individual/Decision Tree Models

Brennan et al. (year) used data from the Health Survey for England (HSE) 2008 to create a representative sample of the general population aged over 16. They utilised a zero-inflated binomial regression model that incorporated variables from HSE (such as sex, work status, and car access) and information from the National Travel Survey (NTS) to estimate the impact of changes in walking or cycling on physical activity levels. This approach allowed for the consideration of activity displacement and provided more accurate net PA gain estimations. However, the model was limited by the lack of subgroup data from the four effectiveness studies and could not perform subgroup analysis. If future effectiveness studies produce better data, the model's flexibility to measure heterogeneity could improve as the base population is nationally representative and includes multiple subgroups (such as ethnicity). In conclusion, the individual level model offers more flexibility in measurement of heterogeneity compared to Markovian cohort models.

Economic evaluations should include a consideration of why the baseline population was chosen, and whether it is representative of the population that is likely to be affected by the intervention. Using a nationally representative survey, or a subset of one, to determine the characteristics of the base population of an individual level model is a simple way of doing this for national policy questions. Individual level models (such as microsimulations and discrete event simulations) make it easier to make probabilities of future events differ by individual characteristics and therefore undertake subgroup analysis.

Theme 3. Limited scope of models to health and narrow outcomes within health

Setting the scope of a model is an important decision which can dramatically change the cost-effectiveness of any intervention (Squires et al., 2016b; Smith & Petticrew, 2010). Setting too narrow a scope risks the creation of a model naïve to the broader implications of the decision, while setting too broad a scope can risk over-complicating the decision problem and allowing un-important factors to influence the decision process. Several recent review papers have recommended the inclusion of non-health benefits of public health interventions within models (Squires *et al.*, 2016a; Smith & Petticrew, 2010; Weatherly *et al*. 2009).

Table 6 shows the health benefits and non-health benefits included in the six models. The models differed in the health conditions they included: Fordham & Barton use a simple QALY health gain estimate, while Brennan et al. include only all-cause mortality benefits and Bending et al., Anokye *et al.*, and Campbell et al. focus on the same three core conditions (CHD, Stroke and Type 2 Diabetes), although Anokye *et al.* supplement this with short-term mental health gains. Love-Koh & Taylor extends the three core conditions to five, by adding breast & colon cancer, and include all-cause mortality similar to Brennan *et al.* The limited number of conditions included within the models is, to some extent, a symptom of the limited quality of the epidemiological evidence base. However, this evidence base is developing rapidly with more research being conducted into the health effects of physical activity, with a particular focus on mental health and a variety of cancers (Rebar et al. 2015; Mammen & Faulkner, 2013; Moore et al., 2016). Utilising this evidence base, particularly for mental health which has higher incidence rates in young people compared to the other PA related conditions, may change the estimates of cost-effectiveness. None of the models included a consideration of the negative health effects positively associated with increased physical activity (e.g. injuries), although Bending et al. (2008) state that including injuries is "*unlikely to significantly impact on model output*" while Anokye, Lord & Fox-Rushby (2012) conclude that "*the evidence on injuries suggests that they are rare (Munro et al. 2004) and not expected to significantly affect results when considered at a population level*".

Table 6 Health and non-health benefits included in each NICE physical activity model.

Model	**Health Benefits**	**Non-Health Benefits**
Fordham & Barton (2008)	All-cause health gain	None
Bending *et al.* (2008)	CHD, Stroke, Diabetes	Reduced absenteeism
Brennan *et al.* (2012)	All-cause mortality	Pollution, Congestion, Environment.
Anokye *et al.* (2012)	CHD, Stroke, Diabetes + ST MH Gain.	None
Campbell *et al.* (2013)	CHD, Stroke, Diabetes	None
Love-Koh & Taylor (2017)	CHD, Stroke, Diabetes, Breast & Colon Cancer, All-cause mortality	None

Whether or not the models included non-health effects of interventions was largely dependent on the topic and perspective taken. For example, Brennan et al.'s walking and cycling model include estimates of the effect of interventions on pollution, congestion, and the environment, while Bending *et al*.'s workplace intervention model includes estimates of the effect of PA interventions on labour market absenteeism. Given that the scope of the two models was 'Local measures to promote walking and cycling as forms of travel or recreation' and 'An Economic Analysis of Workplace Interventions that Promote Physical Activity' the inclusion of these variables is not surprising. However, none of the other four models attempted to estimate any non-health benefits: Anokye *et al.* and Campbell *et al.* do not include any mention of non-health benefits, Love-Koh & Taylor recognise that there may be additional benefits to reducing illness in productive members of society but don't include this in the model, and

Fordham & Barton (2008) state that there is some evidence of "*long-term educational gain or class-room performance of increased physical activity*" but it is not included.

The inclusion of non-health benefits may be increasingly important for public health decision makers who are now more closely aligned to local authorities which have constrained resources and many non-health related objectives (Marks et al. 2013). The development of models which demonstrate effects on other sectors may demonstrate the increased efficiency associated with the reallocation or sharing of resources between sectors. For example, a model of physical activity interventions in schools which demonstrated academic benefits may justify shared funding by Public Health England and the Department of Education. Alternatively, if the benefits of an intervention to one sector (i.e. Health to the Department of Health) outweigh the costs to another sector (i.e. Department of Education) there may be efficiency gains achieved via compensation between departments (Claxton et al., 2007)

Physical activity models should consider including a broad range of conditions, including mental health conditions. Negative consequences of interventions on health should be considered, even if not required. Physical activity models should attempt, when appropriate, to consider the effect of interventions on sectors other than health, especially where inter-departmental transfers could lead to efficient reallocation of resources.

Theme 4. Limited consideration of health inequalities

The Health and Social Care Act of 2012 requires the Secretary of State in England to "*have regard to the need to reduce inequalities between the people of England with respect to the benefits that they can obtain from the health service*". The NHS and PHE typically focus on health outcome disparities, but other organisations such as Sport England may also be concerned with disparities in physical activity. Informed decision making is made possible by presenting decision makers with information regarding the balance between efficiency and equity (Cookson et al., 2017; Johri & Norheim, 2012).

However, economic evaluations of public health interventions do not always include analysis of the effect of interventions on health inequities. It has been argued that this is because many of the methods used to model public health interventions have been borrowed from models of health technologies, in which 'a QALY is a QALY is a QALY' (Weatherly et al. 2009, Cookson *et al.*, 2009; Squires *et al.* 2016b).

None of the six PA models included within this review model the effect of PA interventions on health inequities (as shown in Table 7). In each case limitations in model inputs meant that this was not feasible; Bending *et al.* because the effectiveness studies did not include information on the characteristics of individuals who accessed the services, Fordham & Barton (2008) because their initial population is assumed homogeneous, and Anokye et al. and Campbell et al. don't have a base population with the necessary variables. Brennan *et al.* (2012) briefly consider the effect of the intervention on inequalities, suggesting that walking and cycling interventions have the potential to reduce inequalities since car ownership is positively correlated with socioeconomic status and therefore

improved walking, cycling and public transport reduces inequalities in mobility. However, the authors included no quantitative analysis of intervention effectiveness at reducing inequalities. Love-Koh & Taylor (2017) uses a method developed by Minton *et al.* (2013) to vary the distribution of PA increases in the population, thereby considering the effect of the distribution of PA increases on overall cost effectiveness. However, they provide no analysis of how the interventions change existing inequities in health outcomes within the population studied.

Table 7 Consideration of inequality in the NICE Physical Activity models

Model	**Inequality in Physical Activity Levels**	**Inequality in Health or non-Health Outcomes**
Fordham & Barton (2008)	No	No
Bending *et al.* (2008)	No	No
Brennan *et al.* (2012)	Partial	No
Anokye *et al.* (2012)	No	No
Campbell *et al.* (2013)	No	No
Love-Koh & Taylor (2017)	Yes	No

A recent systematic review of the equity effects of physical activity interventions targeted at children found that only 12% of studies considered differences in intervention effect by socioeconomic status (Love et al. 2017). Given that methods have recently been developed to report the effect of interventions on health equity quantitatively, such as distributional cost-effectiveness analysis (Asaria et al., 2016), it seems timely for such methods to be applied to a physical activity intervention.

When evaluating physical activity interventions, modellers should consider whether to take into account disparities in health outcomes. If trade-offs between efficiency and equity exist, a more in-depth analysis such as distributional cost-effectiveness analysis may be necessary. Some decision makers may also be concerned with inequalities in physical activity for its own sake. By providing such information, decision making can be made more informed.

Theme 5. Limited or no modelling of intervention effect decay or tracking of physical activity over the life course.

The long-term cost-effectiveness of physical activity interventions is almost always dependent on the sustainability of behaviour change over time. This raises two important questions: 1) How is physical activity expected to change over the life-course, both relative to others and absolutely, in the absence of the intervention? And 2) How does the impact of the intervention fade over time? If physical activity

is expected to remain stable without the intervention and the impact of the intervention doesn't diminish, then evaluating the intervention's effect is more straightforward.

Although Love-Koh and Taylor considered changes in physical activity distributions for different age-groups, they did not take individual level variation into account. None of the other five physical activity models considered this aspect, instead assuming that physical activity remains relatively constant throughout a person's life. As such, the sensitivity of the results to this assumption has not been tested. Changes in PA as people age has the potential to impact both efficiency and equity of interventions. Interventions targeted towards groups with more stable PA trajectories may result in greater efficiency. Different levels of stability in PA over the life course for different sub-groups may also predict intervention effectiveness and influence equity of effect. Incorporating this into a model is hard, both from a statistical perspective but also when explaining to stakeholders but understanding it may help identify the optimal ages for delivering interventions.

The treatment of the decay in intervention effectiveness is also undeveloped in the models. Bending *et al.* (2008) simply state that "*In all cases it is assumed that the resulting increase in physical activity is maintained long enough to obtain the health benefits associated with that level of physical activity*" (p.19) while Anokye *et al.* (2011) introduce the first consideration of decay of effect into a NICE PA cost effectiveness model, varying the length of effect between lifetime, 10 years and 1 year, a method replicated by Campbell et al. While the authors argue that the longitudinal evidence base incorporates natural decay over the life-course there is likely to be a decay of the effect of the intervention itself independent of PA decay over the life-course. Brennan *et al.* (2012) improve upon this methodology, creating decay rates, in percent per annum, for their continuous measure of PA. They find that the ICER is very sensitive to decay of effect, and therefore that "*it would be sensible to monitor these factors during the lifetime of any intervention*" (p.20). Love-Koh & Taylor (2013) applies a decay in the first year after the intervention for a fixed proportion of the cohort, though it is unclear what happens after this year. In all cases decay is assumed to be homogenous. Applying different rates of decay for subgroups may help improve the external validity of models and may help to better describe the distribution of outcomes to decision-makers who want to consider both efficiency and equity.

Models evaluating physical activity interventions should take into account the decay of the intervention's impact over time alongside the changes in physical activity levels that would occur over time in its absence. It would be ideal if the models could account for variability in the decay rate among different individuals in the population. If actual data on decay rates is not available, effectiveness evidence should aim to provide an estimate of expected decay that can be used in the models.

Conclusion

Physical activity models developed for NICE improved between 2008 and 2017 as the intervention effectiveness and epidemiological evidence improved. The movement away from PA as a binary

variable, towards continuous measures of physical activity, has increased the precision with which models can estimate gains in population health from interventions which have small but significant impacts on the distribution of activity in the population. The models have developed over time to account for a broader range of diseases alongside all-cause mortality. However, there is relatively little consideration of effects on mental health despite a rapidly growing epidemiological evidence base in this area.

The models differed a lot in their approach to decaying intervention effects, with some models ignoring the problem completely. Two of the models included a consideration of the non-health benefits and costs of PA interventions in sectors including: labour markets, transport, and education. While some of the models included subgroup analysis, there was a limited consideration of inequities in health and physical activity.

The findings of this review have several implications for the development of future physical activity models. Firstly, models can develop the methods of *Brennan et al.* by treating physical activity as a continuous variable and incorporating displacement effects between PA physical activity types, and *Love-Koh & Taylor* by using METs as the continuous measure linking directly to specific diseases and all-cause mortality. Models could be developed which consider more than just the '*big three*' physical activity conditions, including others such as breast and colon cancer, and crucially, mental health conditions. Finally, models can do more to provide decision makers with useful information on inequities, both in health and physical activity.

There are several limitations of this review. Firstly, this review focused on the model structure and techniques, and was not concerned with accompanying evidence which may have been significant in the NICE decision-making process. The focus is also only on physical activity models, and many of the methodological limitations discussed may have been overcome by other public health risk factor models. There may therefore be scope for learning from economic evaluation in other areas of public health, for example tobacco or alcohol models. Lastly, this review only considers the six NICE commissioned physical activity models, and therefore the findings may not apply to economic evaluations undertaken in academic institutions or for other decision makers, which is the focus of the next section.

Physical Activity Models in the Academic Literature

Background

This section of the review focuses on a wide range of cost-effectiveness analyses of interventions aimed at promoting physical activity as a preventative measure. The studies were mostly conducted in western countries, and most of them have been published in peer-reviewed journals or via government or regulator reports. While a full examination of the methods used in every economic evaluation of physical activity is beyond the scope of this thesis, several systematic reviews have already been conducted on the topic. This section highlights the strengths and limitations identified in these reviews and compares key themes to the findings of the review of 6 economic evaluations produced for NICE.

In August 2017, Abu-Omar et al. published a systematic review-of-reviews on the cost-effectiveness of physical activity interventions. The authors searched ten databases and conducted a manual search of grey literature sources between January 2000 and October 2015 for articles which either directly modelled or summarised the results of other health economic evaluations of interventions targeted at increasing physical activity in healthy children, adults, or older people. A total of 18 reviews met the inclusion criteria. While Abu-Omar et al. provided a very brief summary of the modelling method limitations identified in the 18 reviews, this review provides a more in-depth examination of these limitations. This review was originally conducted in 2017-2018 and later updated to include materials published from 2018 to 2023.

Methods

All 18 papers identified by Abu-Omar were downloaded and reviewed in the order shown in Omar-Abu et al., 2017, p.75. Upon second reading, data was extracted including: search strategy, inclusion/exclusion criteria, number of studies, methodology used (e.g. cost-effectiveness, cost-utility). Also extracted were any comments in the text which specified methodological weaknesses or methodological strengths of the economic studies. These could be specific to a single study or more general. These were compiled, and key themes were identified.

Results

The 18 reviews, as identified by Abu-Omar et al. 2017, are summarised in Table 8 below.

Table 8 Studies Included in the review of the academic literature

#	Author, year	Type	Timeframe	No.	Method
1	Laine et al. (2014)	Multiple	to June 2013	14	Cost effectiveness
2	Cavill et al. (2008)	Cycling and Walking Models	to December 2006	16	Multiple
3	Wolfenstetter and Wenig (2010)	Physical Activity for Adults.	to December 2009	15	Cost-utility & Cost-effectiveness
4	GC et al. (2015)	Brief Interventions	to August 2014	13	Multiple
5	Campbell et al. (2015)	Exercise Referral Schemes	October 2009 to May 2013	2	Cost-utility
6	Pavey et al. (2011)	Exercise Referral Schemes	1990 to October 2009	7	Cost-utility & Cost-effectiveness
7	Williams et al. (2007)	Exercise Referral Schemes	to March 2007	4	Cost-utility & Cost-effectiveness
8	Foster et al. (2013)	Remote and Web Interventions	to October 2012	3	Cost-effectiveness
9	Gorden et al. (2007)	Multiple	1995 to 2005	13	Not explicitly stated
10	Windle et al. (2010)	Mental Wellbeing in Old Age	1993 to February 2007	1	Cost-utility analysis
11	Davis et al. (2009)	Falls Prevention	1945 to July 2008	9	Cost Benefit, Cost-Effectiveness, Cost-Utility
12	Balzer et al. (2012)	Falls Prevention	2003 to January 2010	21	Multiple
13	Wu et al. (2011)	Multiple	2000 to June 2008	91	Not explicitly stated
14	Lewis et al. (2010)	Multiple	1995 to September 2010	53	Multiple
15	Muller-Riemenschneider et al. (2009)	All	August 2001 to June 2008	8	Multiple
16	Garrett et al. (2011)	Primary Care Interventions	2002 to 2009	13	Cost-utility & Cost-effectiveness
17	Lehnert et al. (2012)	All	to June 2011	3	Cost-utility analysis
18	Van Dongen et al. (2011)	Worksite interventions	to January 2011	2	Cost Benefit

The 18 reviews included in Abu-Omar et al.'s published study varied in terms of the number of physical activity studies analysed, ranging from 2 to 90, and the type of economic analysis methods used. Some reviews only analysed cost-effectiveness studies while others used a variety of methods. The reviews also differed in their scope, with some focusing on the general population while others targeted specific subgroups or localities. Almost all the reviews highlighted the diversity in the methods used in economic evaluations, which indicate a tailoring of methods for specific decisions but may also indicate a lack of consensus on the best methodological approach. Despite this diversity, several common methodological challenges and limitations emerged from the reviews including inconsistent outcome measures, lack of one way and/or probabilistic sensitivity analysis, lack of consideration of long-term effectiveness, and a narrow perspective. These are each outlined in more detail below.

Differing Outcome Measures

The use of different outcome measures in economic evaluations make comparison between findings difficult. Four of the reviews explicitly refer to the lack of use of QALYs as a problem (Wolfenstetter and Wenig, 2010; Pavey et al., 2011; Vijay CG et al., 2015; Garrett et al., 2011). Wolfenstetter and Wenig (2010) find that of the fifteen studies included in their review "*only three studies used QALYs as an outcome variable, the other studies could not be compared regarding their outcomes*" (p.1638). A similar finding is made by Garrett et al. (2011) who found that due to differences in "*outcome variables between the studies reviewed, it is difficult to draw firm conclusions about which types of interventions are most cost-effective*". It is unclear why cost-utility models with QALYs as the outcome variable is not the standard approach, although Pavey et al. (2011) found that in one case the main limitation of a particular study was "*the inability to convert the findings presented in the form of SF-36 scores into utility scores that might allow for the derivation of QALYs*" (p.55). In response to this heterogeneity Vijay CG et al. (2015) state that what is needed is a single framework, or decision analytic model to identify the most cost-effective interventions.

Lack of sensitivity analysis

The sparsity of sensitivity analysis on uncertain variables is another regular theme, especially in reviews of earlier studies, many of which undertook no formal parametric or structural sensitivity analysis. Van-Dongen et al. (2011) found that "*few studies conducted a sensitivity analysis and hardly any of the studies reported on the uncertainty around their financial return estimates*" (p.1046). However, there has been a general improvement over time, with more recent reviews being less likely to suggest that insufficient sensitivity analysis has been conducted. Lehnert (2010), for example, found that studies included in the review all conducted sensitivity analysis, with most including probabilistic sensitivity analysis.

Lack of long-term effectiveness data

There is a recognition of a lack of long-term effectiveness data, with studies often being too short to reliably estimate long run effects of interventions (Foster et al., 2013; Garrett et al., 11; Pavey et al., 2011). The lack of long-term effectiveness data has led to a variety of strategies used to predict long-term effects, with the most common being a simple assumption about continued effectiveness after 1 year accompanied with sensitivity analysis (Windle et al., 2010). The overreliance on assumptions spurred Pavey et al. (2011) to ask whether the simple decision-analytic approach to modelling is the best method. "*Given that individuals' behaviours may change over time, it may be that a more dynamic approach to modelling the cost-effectiveness of PA is warranted, although once again this may be limited by the available evidence*" (p.57).

Perspective Taken

While few reviews mentioned the perspectives taken by studies as a weakness, it was noted in more than one review and is therefore mentioned here. Muller Riemenschneider et al. (2009) and Garrett et al. (2011) note the many different perspectives. Lehnert et al. (2012) note that many studies did not take a full societal perspective and missed many potential spillover effects, particularly in the education sector. Van Dorgen et al. (2011) found that absenteeism was included in some cases, although presenteeism was harder to include in the models. While it is important to recognise that economic models are created for very different purposes and the use of a different perspective is not necessarily a limitation, the heterogeneity of the perspectives taken does limit comparison between evaluations of interventions. Encouraging future cost-effectiveness studies to report results from a consistent set of perspectives, such as a full societal perspective or an NHS perspective only, could enhance future comparisons.

Discussion & Conclusion

The 2017 review by Abu-Omar et al. identified several limitations in physical activity (PA) models. Most of these limitations were also noted in this review-of-reviews. One limitation identified by Abu-Omar et al. of the literature in aggregate, but not identified in this review of individual studies, was that individual-level interventions are more frequently evaluated compared to population-level interventions.

Since this review aimed to provide a more detailed analysis of the methodological limitations and challenges identified by the reviews, there were several limitations which were identified which were not included in the Abu-Omar paper or were not discussed in depth. Firstly, several reviews, including Laine et al. (2014) estimated the cost per unit increase in physical activity, generally measured in METs or minutes. This is a simpler task than attempting to estimate a cost per QALY, but this approach is fundamentally flawed in public health models, as explained in the steps below:

1) There is a widely accepted curvilinear relationship between physical activity and health outcomes (Kyu et al. 2016). Therefore, the benefits of additional physical activity differ substantially between individuals depending on baseline activity.

2) There are complex biological, psychological, social, environmental, and regional (Bauman et al., 2012) reasons why some people are more-or-less active than others. Increasing the physical activity levels of those who are least active is often the most difficult.

3) Cost-effectiveness models of physical activity which evaluate average gains in physical activity in a population may result in allocation of resources **away from** interventions which generate the largest health gain, towards those which have the largest physical activity gain.

If the aim of the decision-maker is to maximise health outcomes from physical activity, models should adopt a cost per QALY approach, or should attempt to report how interventions may change physical activity inequality.

Decision makers who want to compare the cost-effectiveness of physical activity (PA) interventions with other types of interventions, such as weight loss, dietary interventions, or pharmacological interventions, would benefit from high-quality models that report cost per QALY as an outcome - such as those of Roux et al. (2008) and highlighted in the Lehnert et al. review. These central estimates should incorporate uncertainty and be accompanied by one way sensitivity analysis for key parameters and structural assumptions. It is crucial that modellers find ways to overcome the limitations of short RCTs, potentially through increased use of expert elicitation, a method that is currently underutilised in PA models. Ultimately, reaching a consensus on the most appropriate methods of modelling the long-term effects of PA interventions would enhance transparency and confidence in the models.

Update of review from 2018 to 2023

Methods

The same process described in the review of NICE guidance in 2018 was undertaken in February 2023. No further economic modelling studies were identified.

To identify whether other systematic reviews of the strengths and limitations of evaluations of physical activity models had been undertaken since 2018 the following approach was taken: (1) All 116 papers citing Abu-Omar to 1st Jan 2023 were assessed for suitability. (2) A Google Scholar search was undertaken using the combined combinations of key terms ‘cost-effectiveness’, ‘physical activity’ ‘modelling’. Paper titles and abstracts were scanned with the requirement that the papers were written in English, and the focus of the paper was on the modelling methods used to estimate the cost-

effectiveness of interventions which aim to increase physical activity. The references of all papers identified were scanned in a snowball approach. Papers were read in chronological order.

Results

No papers from the 116 papers citing Abu-Omar et al. 2017 met the criteria. Three papers were identified by the Google Scholar search (Hazel et al. 2021; Candio et al., 2020; Cochrane et al. 2019). The references of these papers and the papers citing them were reviewed, but none were deemed to match the criteria.

Cochrane et al. (2019) finds several challenges for health economic models of physical activity interventions: attributing effects given short follow up, inconsistencies in types of measured outcomes, inconsistent perspectives chosen and a lack of consideration of equity. They find that overall, the lack of long-term follow up and limitations in methods used to estimate long-term effects of interventions mean that "*Curative interventions that rescue people from very poor health to better health* [in the short term] *will continue to be favoured, even if they are less cost-effective overall.*" (p163) since "*Any future reduction in incidence of NCD and premature mortality, attributable to physical activity and sedentary behaviour interventions, is unlikely to manifest until decades after the intervention has taken place*". The authors argue that while the incorporation of long-term effects is hampered by availability of data, it is the modeller's "*time and skills*" that are more often the limiting factor, encouraging the development of "*novel modelling skills*" and citing modelling studies that have included estimates of long-term costs and benefits as best practice (Campbell et al. 2015; Anokye et al. 2012). Like the review described above, the authors also identify that the diverse range of perspectives, outcomes, and intersectoral effects make comparisons between studies difficult.

Candio et al. (2020) aimed to "*provide an overview and critique of modelling approaches and key structural assumptions used in applied studies to estimate the impact of physical activity on population health.*" (p. 1155) and come to similar conclusions to my review above conducted to 2018. The authors highlight two core limitations of models as being particularly important: (1) "*assumption of no decay of intervention effect over time is unrealistic*", (2) "*health equity concerns have not been incorporated into the models*" and suggest that (3) "*development of a reference model could help reduce variability in modelling approaches*".

The first of these has been discussed above, the authors found that "*The majority of models (15/25) assumed implicitly or explicitly that the intervention effect would not decay after the intervention ended (i.e. beyond follow-up assessment period)*" and that "*only a minority of models explored the impact of variations to these base-case structural assumptions*" (p. 1157). Baseline PA was also assumed to be stable in 15 of 25 studies with horizons > 30 years, meaning long term trajectories of PA had not been considered.

For the second, while many studies did run sensitivity analysis for specific groups in society, "*None of the reviewed economic models incorporated concerns relating to the distributional impact of the intervention formally into the economic evaluation*" (p. 1158).

The third issue is not a limitation of any given model, but a reflection that heterogeneity in modelling methods may indicate that suboptimal modelling techniques are being used. The authors state that:

"*While acknowledging that the trade-off between simplicity and internal validity still represents an unresolved challenge for modellers, this literature is predominantly characterised by modelling approaches that may not adequately address the complexities associated with the PA behaviour – population health process they were intended to represent*." (p1158).

The author suggests five minimum modelling standards, summarised here:

1) Modelling of downstream disease risks based on epidemiological evidence.
2) Accommodating for the dynamics of PA, using natural trends as the baseline for comparison.
3) Making explicit assumptions around lags and decay of intervention effect and including scenario analysis.
4) Accounting for differences in effects, risks and outcomes due to the characteristics of individuals.
5) Better incorporation of uncertainty with a more structured assessment of structural assumptions.

These standards serve as a useful checklist for future model development, and also help to develop novel modelling skills as recommended by Cochrane et al. (2019).

The most recent paper, by Hazel et al. (2021), reviewed the methods used to estimate the cost-effectiveness of behaviour change communication apps. The review was limited by a small number of studies (n = 6), and the limited methods utilised by those studies. One challenge flowed from a short follow up of trials, meaning that it was hard to identify any impact in terms of QALYs (the predominant outcome measure). As a result, there were some attempts to either value intermediate outcomes or to use intermediate outcomes as links to long term outcomes which can be measured using QALYs (as recommended by Candio et al. 2020). The former results in a plethora of outcomes that make comparison between studies hard (as identified by Cochrane et al. 2019), whereas the latter results in considerable uncertainty and methodological challenges in long term extrapolation.

Discussion & Conclusion

Many of the limitations identified in the review of NICE studies and the previous literature remain limitations in the three studies identified in this update. In particular, Candio et al. 2020 find many of the same limitations and make many of the same recommendations, albeit with less focus specifically on models developed for NICE and with access to some newer publications.

Implications and main findings of review

The review of NICE guidelines and related literature revealed several shortcomings in current models for increasing population physical activity. The most significant limitations are:

1) The need for a single measure of physical activity to better align surveillance, epidemiological, and health economic modelling studies.
2) The need to consider physical activity as a continuous variable to reflect non-linear benefits and small changes in physical activity distributions at a population level.
3) The need to model a broader range of diseases.
4) The need for modelling methods to estimate the long-term costs and benefits of interventions, and therefore to include estimates of long-term physical activity trajectories in the absence of intervention.
5) The importance of estimating the decay in the effect of interventions over time.
6) The need to consider differences in outcomes among subgroups, and distributional effects across society.
7) The need for probabilistic and sensitivity analysis of key parameters and structural assumptions in models.
8) The importance of fluctuations in physical activity over the lifespan due to seasonality and key life events.
9) The need to better incorporate competing risks and feedback loops in models, using calibration methods to account for disease dependencies.

All evaluations took the form of a written report, with varying levels of sensitivity analysis, and none referred to a user-interface or modelling tool which could be used by decision makers (although this does not discount the probability of there being one). In addition, none of the models described were made open source, and therefore it was difficult to assess the overall usability of the models.

The next sections of this thesis aim to further investigate how to adapt existing models and build new ones to improve usability and external validity and navigate trade-offs where they exist.

Part 2 - Adapting an existing model - case study of the WHO HEAT model

The aphorism *'All models are wrong, but some are useful*,' commonly attributed to George Box, is accepted in the field of health economics. However, there is often a trade-off between usability and external validity in health economic models. On the one hand, a model with high external validity may be more reliable for making decisions and predicting outcomes, but it may also be more complex and difficult to use, requiring specialised training or expertise to understand and apply. On the other hand, a model with high usability may be easier to understand and apply, but it may also have lower external validity, leading to potentially flawed or misleading conclusions. Striking the right balance between usability and external validity is an important consideration in the development and application of health economic models. As noted in a review of the scientific response to the pandemic, "*A model that focuses on the key parameters is a lot more useful than a more complicated one that tries to bring in everything*" (House of Commons, 2021; p.43).

This section of the thesis uses the Health Economic Assessment Tool (HEAT) for walking and cycling as an example of a model which is particularly simple but has proven to be incredibly *useful* (Kahlmeier et al., 2020). The HEAT is a tool developed by the World Health Organization (WHO) to estimate the health and economic benefits of promoting active transportation, such as walking and cycling. HEAT has been used in a number of countries around the world to assess the potential benefits of investing in infrastructure and programs to promote walking and cycling (Brown et al., 2016; Gao et al. 2017; Genter et al. 2018). In the UK, HEAT has been used to assess the potential benefits of investing in walking and cycling infrastructure in England, Scotland and Ireland (Deenihan & Caulfield, 2014; Sustrans, 2014).

The tool allows policymakers and planners to calculate the potential cost-effectiveness of investments in infrastructure and programs to promote active transportation. The HEAT uses a mathematical model, programmed in the R software environment and with a web-based user-interface (since Version 4.0), to estimate the number of deaths that could be prevented, and the economic value of those lives saved based on the amount of walking and cycling that is expected to result from a particular intervention. The tool also estimates the reduction in greenhouse gas emissions and air pollution, and increased crash risk associated with active transportation.

The focus of this work has been on the physical activity module of the HEAT. To use the physical activity module, users input data about their population, including information about age, gender, and current levels of physical activity (active/inactive). Users also need to input data about the intervention, including information about the expected increase in physical activity that it is likely to produce (in minutes of walking/cycling). The tool uses this data to estimate the potential deaths averted, assuming a linear

relationship between physical activity and all-cause mortality with a maximum relative risk reduction (Kahlmeier et al. 2017). It then calculates the associated societal value of averted deaths using the Value of a Statistical Life (VSL) (Viscusi & Aldy, 2003).

The three chapters in this section of the thesis focus on improving the *external-validity* of the HEAT for walking and cycling, while also considering the trade-offs between *usability* and *external-validity*. The first chapter, published as Smith et al. 2021a, investigates the effect on net monetary benefit estimates of using "life-years" gained as a measure, rather than "deaths averted". The results suggest that this approach may improve the external validity of estimates when the intervention population age distribution differs from that of the general population but may also make it more difficult to explain to users. In the second chapter, published as Smith et al. 2022c, the current linear dose-response relationship between physical activity and mortality is compared to a non-linear relationship. The study finds that the use of a non-linear relationship leads to significantly different estimates for populations that are particularly inactive or active. The third chapter explores the broader implications and trade-offs between the usability and external validity of health economic models.

These studies were conducted during my research placement with the University of Zurich's World Health Organization Physical Activity Unit, and the findings and views expressed are my own and do not reflect those of the unit or the WHO-Europe HEAT team.

Chapter 3. HEAT - The Value of the Statistical Life Year

This chapter contains a publication which reports the results of the first case study using the HEAT in which the net monetary benefit associated with reductions in mortality rates are compared with a valuation method which uses the value of a statistical life (current approach) and the value of a statistical life year (VSLY) (proposed approach).

https://www.sciencedirect.com/science/article/pii/S0033350621001256

The paper in the chapter was written with 5 co-authors: Chloe Thomas, Hazel Squires, Thomas Götschi, Sonja Kahlmeier and Elizabeth Goyder. Robert Smith was first author and corresponding author, leading the conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualisation, and writing, reviewing, and editing the article. Chloe Thomas, Hazel Squires and Elizabeth Goyder contributed to supervision and project administration. Thomas Gotschi and Sonja Kahlmeier contributed to data and code provision for the current methodology. All authors contributed to the review and editing of the final manuscript. This work, or variations thereof, does not form a part of any other PhD thesis.

Public Health 194 (2021) 263–269

Contents lists available at ScienceDirect

Public Health

journal homepage: www.elsevier.com/locate/puhe

Original Research

The price of precision: trade-offs between usability and validity in the World Health Organization Health Economic Assessment Tool for walking and cycling

R. Smith [a, *], C. Thomas [a], H. Squires [a], T. Götschi [b, d], S. Kahlmeier [b, c], E. Goyder [a]

[a] School of Health and Related Research, University of Sheffield, Regents Court, Sheffield, S1 4DA, UK
[b] University of Zurich, Epidemiology, Biostatistics and Prevention Institute (EBPI), Switzerland
[c] Swiss Distance University of Applied Science, Department of Health, Switzerland
[d] School of Planning, Public Policy and Management, University of Oregon, Eugene, USA

ARTICLE INFO

Article history:
Received 17 November 2020
Accepted 16 March 2021
Available online 14 May 2021

Keywords:
HEAT
Physical activity
VSL
VSLY
Health Impact Assessment

ABSTRACT

Objectives: The widely used World Health Organization (WHO) Health Economic Assessment Tool (HEAT) for walking and cycling quantifies health impacts in terms of premature deaths avoided or caused as a result of changes in active transport. This article attempts to assess the effect of incorporating 'life-years' as an impact measure to increase the precision of the model and assess the effect on the tool's usability.

Study design: This article is a methods paper, using simulation to estimate the effect of a methodological change to the HEAT 4.2 physical activity module.

Methods: We use the widely used WHO HEAT for walking and cycling as a case study. HEAT currently quantifies health impacts in terms of premature deaths avoided or caused as a result of changes in active transport. We assess the effect of incorporating "duration of life gained" as an impact measure to increase the precision of the model without substantially affecting usability or increasing data requirements.

Results: Compared with the existing tool (HEAT version 4.2), which values premature deaths avoided, estimates derived by valuing life-years gained are more sensitive to the age of the population affected by an intervention, with results for older and younger age groups being markedly different between the two methods. This is likely to improve the precision of the tool, especially where it is applied to interventions that affect age groups differentially. The life-years method requires additional background data (obtained and used in this analysis) and minimal additional user inputs; however, this may also make the tool harder to explain to users.

Conclusions: Methodological improvements in the precision of widely used tools, such as the HEAT, may also inadvertently reduce their practical usability. It is therefore important to consider the overall impact on the tool's value to stakeholders and explore ways of mitigating potential reductions in usability.

Introduction

There has been an increasing awareness of the need to incorporate Health in All Policies (HiAP) to ensure that non–health government agencies work in partnership to incorporate considerations of health and well-being when developing policy.[1] One simple way in which HiAP is often facilitated is through quantitative Health Impact Assessments (HIA), simple statistical models of the world, which aim to quantify the costs and benefits of interventions.[2,3] To make HIA easier and cheaper to implement, online tools have been developed, which allow stakeholders to undertake their own HIA.[4,5]

The WHO-Europe's Health Economic Assessment Tool (HEAT 4.2) is an example of a widely used HIA tool designed specifically for a HiAP purpose,[1] allowing decision-makers in the transport sector to incorporate the health implications of walking and cycling

* Corresponding author. 30 Regents Court, Sheffield, S1 4DA, UK. Tel.: 07920746932 (mobile)
E-mail address: rasmith3@sheffield.ac.uk (R. Smith).

https://doi.org/10.1016/j.puhe.2021.03.016

into economic appraisals.[6] The tool has been used directly by public sector decision-makers in different locations, including Kuopio (Finland), Parnu (Estonia), Brighton & Hove (UK), Modena (Italy), and Viana do Castelo (Portugal), and by academics in a number of published studies over the past two decades.[7,8] One of the reasons why the HEAT has been so popular is that it is simple and easy to use, as one of the core principles of the HEAT is to be "as user-friendly as possible".[6]

The HEAT 4.2 has four modules: physical activity, air pollution, crash risk, and carbon emissions.[6] The physical activity module generally accounts for most of the estimated intervention effect.[4,9] Within the physical activity module, the estimated net mortality risk change is valued using the Value of a Statistical Life (VSL), an estimate of the societal willingness to pay for a reduction of one statistical fatality.[10] The measure is commonly used in transport planning.[11,12]

Previous studies have compared the results derived by the HEAT with other HIA tools, such as the Integrated Transport and Health Impact Modelling Tool and Dynamic Modelling for Health Impact Assessment.[13,14] Other studies have assessed the effect of the method used to aggregate benefits within HEAT.[15] However, these comparisons have focused on the effect of the shape of the dose–response relationship between physical activity and health outcomes[13] and the choice of a static vs dynamic modeling methodology.[14] To the best of our knowledge, there are no published studies of the effect of the health valuation method, the valuation of lives saved vs life-years gained, on the results of Health Impact Assessment tools for walking and cycling or physical activity. This paper attempts to fill that gap in the literature.

The VSLY represents society's willingness to pay for reductions in fatality risk, which result in an additional statistical life-year. When using the VSLY reductions in fatality risks, younger populations, with greater expected life-years remaining, are valued more highly than reductions in fatality risks for older populations. When the population affected by a policy is representative of society, valuing premature deaths averted using the VSL and life-years saved using the VSLY are likely, conceptually, to yield similar results. However, when the population is not representative, in terms of age, the two approaches are likely to yield very different results. Attempting to value policies in response to the COVID-19 pandemic made this particularly apparent: multiplying the number of premature deaths averted by the VSL resulted in much higher values than multiplying expected life-years saved by the VSLY since COVID-19 related mortality rates rise super-linearly with age.[16,17] In this article, we argue that the same holds for the HEAT: multiplying premature deaths averted from walking and cycling interventions by VSL is likely to yield different results than multiplying life-years saved by the VSLY if the distribution of age in the intervention group does not match the age distribution implicit in the selected HEAT age group.

We begin by using a simple algorithm to derive estimates of VSLY from the VSL values used by the HEAT. We then compare the results, for the physical activity module of the HEAT, for six hypothetical scenarios using both the VSL and VSLY methods. We focus on how a relatively simple HIA tool, the HEAT, could be adapted to better reflect the age distribution within the active travel population. We also discuss the potential implications of these adaptations on the tool's usability, a core principle of the HEAT,[6] and suggest means by which the tool could remain easy to use.

All data and code (in R software environment) is provided in an open access online repository (https://anonymous.4open.science/r/b1ac653f-7e70-43ab-870c-f3ccc4d63914/).

Methods

Data and measures

This study relies on data used in the HEAT 4.2 and previously described in a study by Kahlmeier et al.,[6] that is, WHO country names, country ISO3 codes, VSL estimates based on the OECD Recommendations on Mortality Risk Valuation in Environment, Health and Transport Policies,[12] and dose–response relationships between walking and cycling and mortality from a study by Kelly et al.[18] This study also makes use of two additional data sets: population estimates and life tables for 2017 from a study by Dicker et al.[19] Table A1 in the supplementary material shows a full list of the variables used in the analysis.

Study design

This paper is a methods paper, using simulation to estimate the effect of a methodological change to the HEAT 4.2 physical activity module.

Analysis

First, we estimate, for each of the 51 WHO European Region countries included in the HEAT tool, the VSLY (in 2015 Euros). We then go on to compare the societal value of premature deaths averted for six scenarios when using the VSLY method, the current HEAT method for the full adult range (VSL-1), stratified by younger vs older adults (VSL-2), and the use of VSL using individual age mortality risks (VSL-55).

Estimating the value of a statistical life-year

The VSL estimate used in the HEAT model is based on a meta-analysis of stated preference studies,[12] in which individuals were asked how much they were willing to pay for a small reduction in mortality risk. The estimates vary considerably between countries, ranging from approximately EUR 143,000 in Tajikistan to almost EUR 7m (2015 values) in Luxembourg. The mean age of participants within the studies in HEAT countries was 50 years. By making the assumptions that (1) the VSL at the age of elicitation is the value derived from future life-years until death and (2) all years are valued equally, it is possible to estimate the VSLY using the equation below. The equation inverts the equations used to calculate the VSL in Annex 1.A1 of the OECD report published in 2012.[12]

$$VSLY = \frac{VSL_{50}}{\sum_{i=50}^{109}\prod_{a=50}^{i} Pr(S)_a \times \frac{1}{(1+r)^{a-50}}} \quad (1)$$

The VSLY is equal to the VSL at age 50 years divided by the discounted expected life-years remaining between age 50 and 109 years, the maximum age in our data. The discounted expected life-years remaining is calculated for each age a, using the probability of survival, Pr(S), to the next birthday, as well as the annual discount rate, r. The VSLY for a country is greater where VSL is greater, annual survival probabilities from 50 to 109 years are lower, or if the discount rate is greater.

The Pr(S) estimates were derived from the Global Burden of Disease Estimates[19] and validated against the UN World Population Prospects life tables.[20] The discount rate, r, was set to zero within this analysis for simplicity because different nations use different discount rates in decision-making. The discounted life-years remaining at each age were validated against the yll package in R.[21]

Estimating monetary benefit using the VSLY

The VSLY method estimates the value of premature deaths averted by (1) estimating the relative risk associated with an intervention, given increases in walking and cycling using a linear dose–response function from;[18] (2) estimating discounted life-years saved, given the relative risk, population age distribution, and baseline mortality rates by age; and (3) multiplying the estimated discounted life-years saved by the VSLY estimate.

The equation is shown below:

$$MB = dLYS \times VSLY \tag{2}$$

Discounted life-years saved (dLYS) can be estimated by multiplying the absolute difference in the relative risk of death (ADRR), estimated using a relative risk function from a study by Kelly et al. (2014), by the age-specific mortality rates MR_i to estimate the effect of an intervention on mortality for the population in each age group pop_i. These changes are then multiplied by discounted expected life-years remaining dLYR_i (itself estimated from Global Burden of Disease life tables) for each age group to give overall discounted life-years saved.

As the absolute difference in relative risk is independent of age, it can be factorized, giving Equation 3 (below) in the case of an intervention affecting 20- to 74-year-olds.

$$dLYS = \Delta RR \times \sum_{i=20}^{74} MR_i \times dLYR_i \times pop_i \tag{3}$$

Inputting this back into our original equation gives:

$$MB = VSLY \times \Delta RR \times \sum_{i=20}^{74} MR_i \times dLYR_i \times pop_i \tag{4}$$

where i has 55 values representing each age from 20 to 74 years.

Note that both VSLY and ADRR are constants while mortality rate, discounted life-years remaining, and population vary with age.

This equation is not substantially more complex than the existing HEAT method (in Equation 5 below), in which monetary benefit is the VSL multiplied by the absolute difference in relative risk associated with an intervention, age group mortality risk, and the number affected.

$$MB = VSL \times \Delta RR \times MR_{20-74} \times pop_{20-74} \tag{5}$$

Comparing four methods for six hypothetical scenarios

To compare the proposed VSLY model with the current HEAT models, we estimate the annual, per capita monetary benefit using four different methods: (1) VSL-1 refers to the current HEAT model with a single mortality rate for the entire population aged 20–74 years, (2) VSL-2 uses the current HEAT model with two mortality rates based on weighted population means (walking: 20–44 and 45–74; cycling: 20–44 and 45–64), (3) VSL-55 uses the existing HEAT model methodology (valuing premature deaths averted using the VSL) but with separate mortality risk estimates for each age from 20 to 74 years, and finally, (4) the VSLY model described previously, using individual ages as in (3) but valuing life-years saved using the VSLY estimates derived earlier. In all cases, the discount rate was set to zero for ease of comparison. We use the four methods to estimate the value of six hypothetical scenarios, three for walking and three for cycling, as shown in Table 1 alongside results for France.

Results

There is considerable heterogeneity in the VSLY estimates of WHO-Europe countries, ranging from EUR 5828 in Kyrgyzstan to EUR 216,838 in Luxembourg, with higher values in western Europe than in eastern Europe. A full table of the VSLY estimates derived are provided in the supplementary material in Table A2 and are broadly aligned with previous estimates of societal willingness to pay for a statistical life-year.[22]

In the first simple scenario, an extra 10-min walking per week for every person aged 20–74 years, the VSLY method results in approximately 25% lower estimated benefits than VSL-1 or VSL-2 (current method with one or two age groups). The effect is not because of more precise mortality rate estimates; the VSL method applied to a population categorized in 1-year age bands (VSL55) results in the same estimates to the VSL model with one and two groups (VSL-1 and VSL-2). Rather, the different estimates for the VSLY are due to assigning our estimates of life-years remaining to each prevented premature death. A full set of results are available in the supplementary material: Table A3 for the three walking scenarios (Scenarios 1, 2, and 3) and Table A4 for the three cycling scenarios (Scenarios 2, 4, and 6).

Fig. 1 displays the results from Scenario 1 graphically for all 51 countries. The current 'best' HEAT method, the VSL with two age groups (VSL-2), is shown on the x-axis as the reference method, and all other methods are depicted in a color-coded scatter plot with a 45-degree line used to depict equity. As these assessments cover the entire HEAT age range (20–74 years), the VSL-1 and VSL-55 estimates are identical to the VSL-2 estimates and therefore lie (jittered) on the 45-degree line. The monetary benefits estimated by the VSLY (blue) are around one-third lower than those estimated by the current VSL-2 model (black line). This is because those with the greatest mortality rates (older people) also have the lowest discounted life-years remaining, thereby reducing the effect that older people have on the mean.

Fig. 1 shows the estimates generated by increased activity in the population aged 20–74 years. However, this masks differences in estimates for the two current HEAT age groups (20–44 and 45–74 years). Fig. 2 depicts the estimates generated by stratifying the analysis to the population aged 20–44 years (left) and 45–74 years (right). In both cases, the VSL55 (green) estimates are equal to the VSL-2 estimates. The VSL-1 (red) method results in higher values when restricting the analysis to younger people and lower values for older people. The VSLY (blue) estimates tend to be greater than that of the VSL-2 in younger people and lower in older people because younger populations have more expected life-years remaining.

Because there are clear differences in the values generated by different methods, and these differences vary between older and younger populations, we also looked at how the valuation methods differ over the life course in an exemplar country. Fig. 3 below shows a comparison of annual monetary benefits per capita (2017 Euro) associated with 10 min/week of additional walking, for each individual age from 20 to 74 years for the Latvian population using the four different models: VSL-1 (red), VSL-2 (black), VSL-55 (green), and VSLY (blue).

The VSL-1 method generates the same results regardless of age, the VSL-2 method generates different results for the population aged 20–44 years to those aged 45–74 years, and the VSLY (blue) and VSL-55 (green) results are similar until around age 55 years, with monetary benefit increasing as age, and therefore, mortality rates increase. However, the VSLY model does not increase as quickly with age because life-years remaining are falling with age also—this is particularly stark from age 60 years onwards.

Finally, it is interesting to observe the differences in results between countries when using the VSLY methods. Fig. 4 shows the estimated per capita annual monetary benefit of an additional 10 min of walking per week per person aged 20–74 years for the HEAT countries on a choropleth map. There are large differences in

Table 1
Monetary benefit estimates for France for each of the six scenarios using the VSL method with two age groups and the VSLY method with individual ages (*assumes scenario population is representative of the general population within that age range*).

Scenario	VSL method result (two groups) in 2017 EUR	VSLY method result in 2017 EUR
Population aged between 20 and 74 do an additional 10 min of walking per week.	86.56	63.75
Population aged between 20 and 64 do an additional 10 min of cycling per week.	77.85	72.5
Population aged between 20 and 44 do an additional 10 min of walking per week.	15.11	21.73
Population aged between 20 and 44 do an additional 10 min of cycling per week.	22.27	32.03
Population aged between 45 and 74 do an additional 10 min of walking per week.	147.27	99.45
Population aged between 45 and 64 do an additional 10 min of cycling per week.	143.42	120.26

VSL, value of statistical life; VSLY, value of statistical life-year.

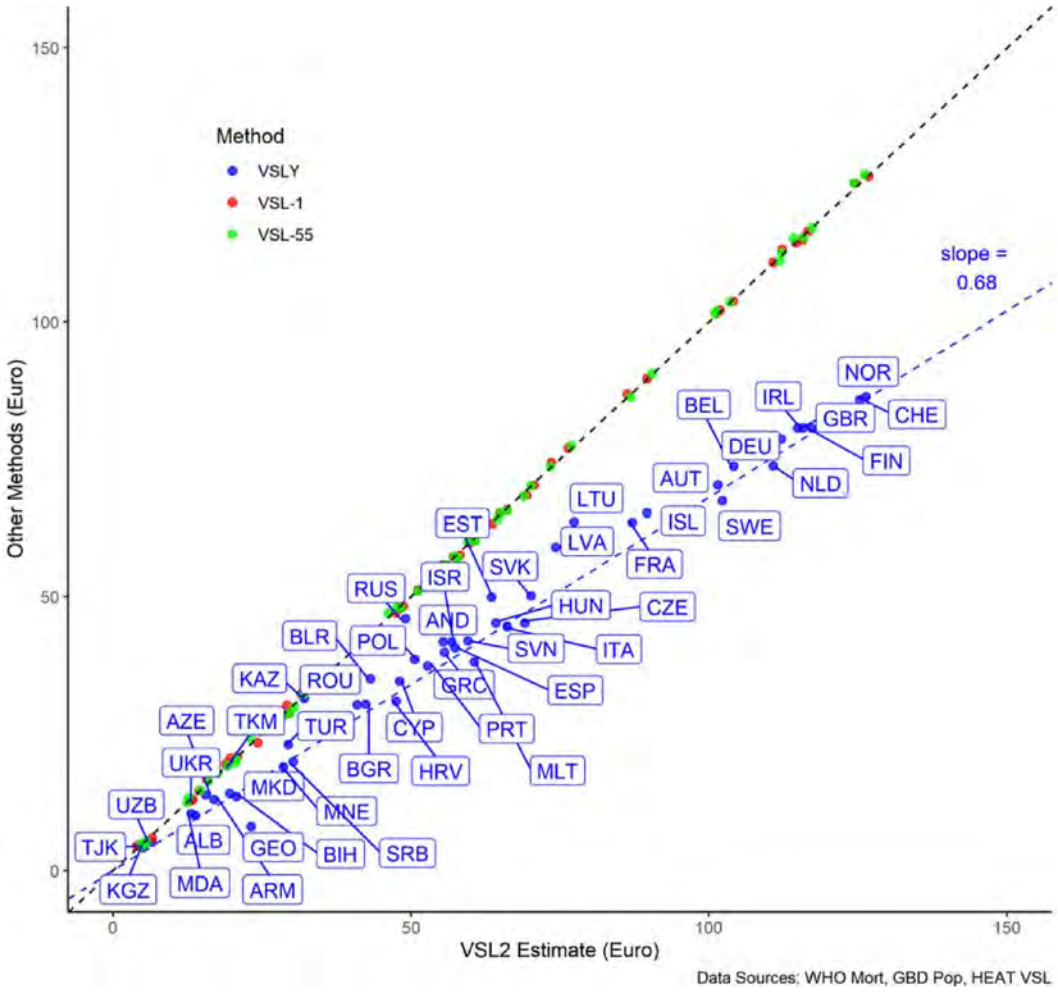


Fig. 1. Estimated annual monetary benefit per capita (in 2017 Euro) in scenario 1, comparing alternative methods to *VSL-2*. VSL, value of statistical life; VSLY, value of statistical life-year; VSL-1, VSL for full adult age range; VSL-2, VSL stratified by younger vs older adults; VSL-55, VSL using individual age mortality risks.

estimated monetary benefit per capita between HEAT countries, with estimated monetary benefit ranging from EUR 4.52 in Tajikistan to EUR 117.13 in Luxembourg.

Discussion

This study is the first to compare the effect of the valuation method used to value averted premature deaths in a Health Impact Assessment tool for physical activity. It uses the WHO HEAT 4.2 for walking and cycling as a case study to compare the estimates of the value of active transport using two different methods: the Value of Statistical Life and the Value of Statistical Life-Year. We show that the VSLY approach generates lower estimates and is more sensitive to differences in the age of the affected population than the VSL with two age groups (VSL-2). However, this comes with a trade-off: although the use of the VSLY may be more accurate, there are additional data requirements of the user. As the minimal data entry requirements of HEAT 4.2 have shown to be a main barrier to wider use of the HEAT, this potential additional user burden warrants serious consideration.

Our findings align with those of previous studies, for example, the work of Robinson et al.,[16] which found that estimates using the VSLY method result in lower valuations of interventions to reduce

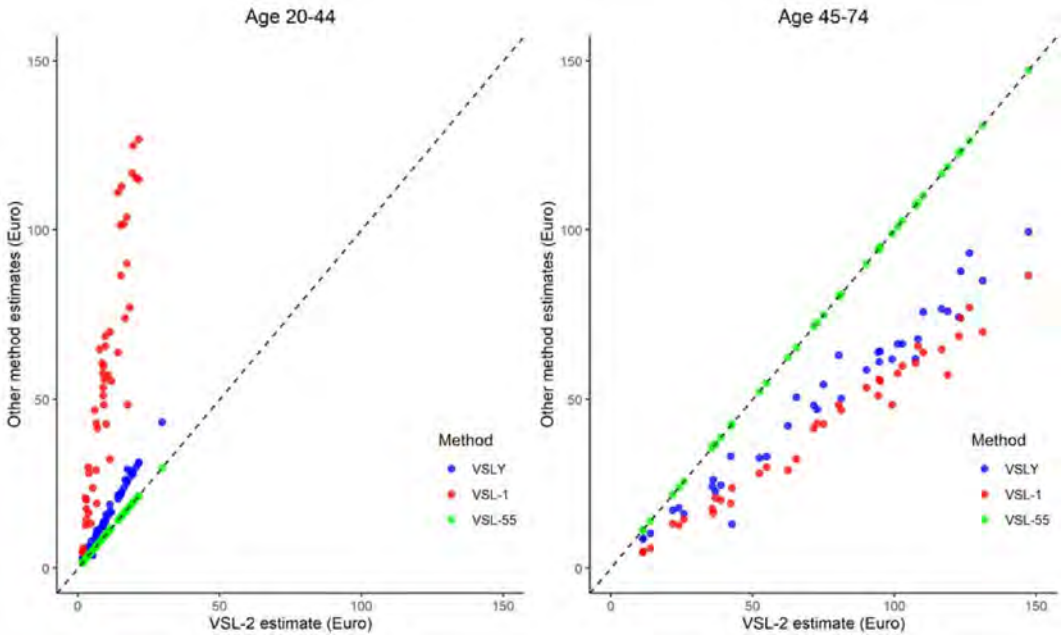


Fig. 2. Estimated annual monetary benefit (in 2017 Euro) per capita from 10-min additional weekly walking using country-specific population age distributions from 20 to 44 years (left) and 45–74 years (right), *VSLY* vs current HEAT models. VSL, value of statistical life; VSLY, value of statistical life-year; VSL-1, VSL for full adult age range; VSL-2, VSL stratified by younger vs older adults; VSL-55, VSL using individual age mortality risks.

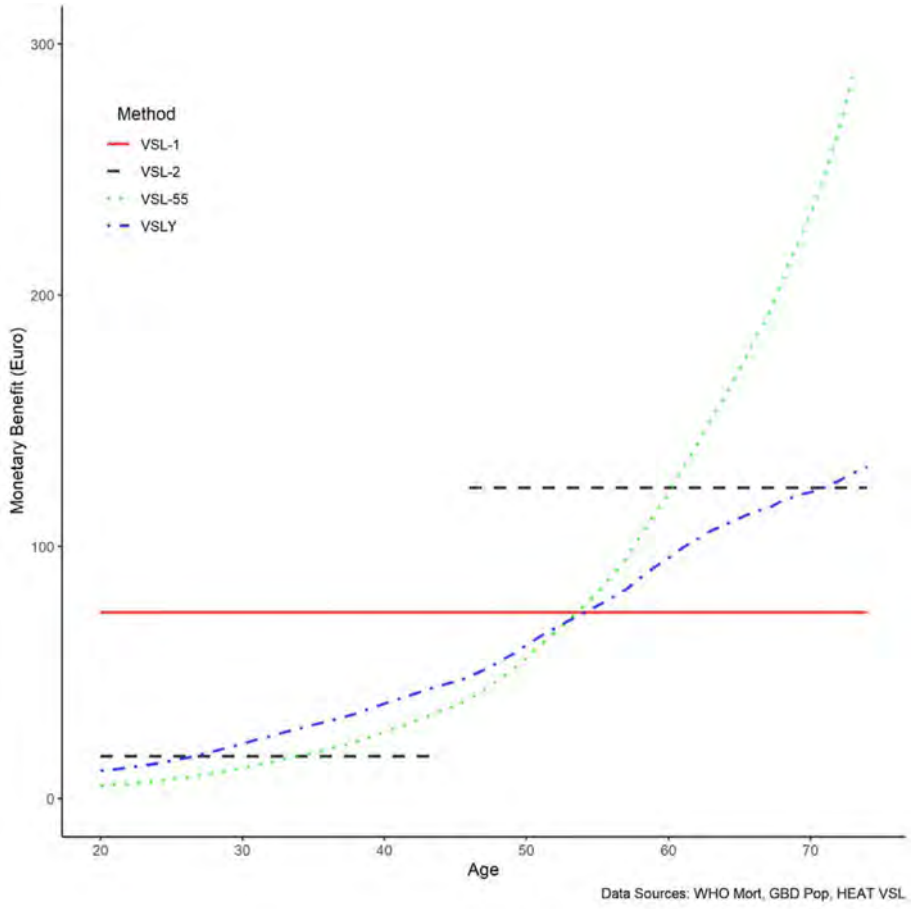


Fig. 3. Annual monetary benefit per capita (in 2017 Euro) from 10-min additional weekly walking for each age of Latvian population, using each method. VSL, value of statistical life; VSLY, value of statistical life-year; VSL-1, VSL for full adult age range; VSL-2, VSL stratified by younger vs older adults; VSL-55, VSL using individual age mortality risks.

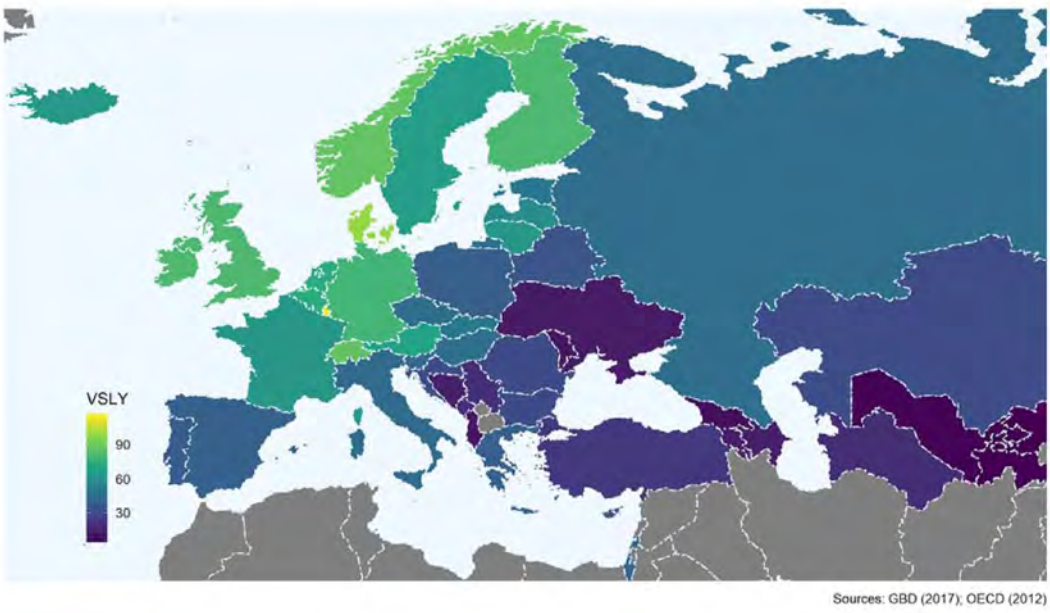


Fig. 4. Map of estimated per capita annual monetary benefit (2017 Euro) of an additional 10 min of weekly walking per person aged 20–74 years for 51 HEAT countries. HEAT, Health Economic Assessment Tool; VSLY, value of statistical life-year.

COVID-19 deaths, primarily from older populations. However, this is the first study that has explicitly analyzed the significance of these methodological decisions for an HIA tool. It is also the first to critique the valuation methods in the physical activity module of the WHO HEAT for walking and cycling. We offer a simple enhancement to the current HEAT physical activity module, which remains within the framework used by transport planners but incorporates the duration of life.

Differences in the estimates using VSL and VSLY methods provoke normative questions about the valuation of premature mortality. The VSL values mortality risk equally irrespective of age, thereby valuing a year of expected life more highly for older persons. On the other hand, the VSLY assigns a constant value to *s* life-year, but, as a result, values mortality risk reduction in younger persons more highly.[17] Transport economics typically uses the former, health economics the latter (and includes quality of life). As an HIA tool used widely in transport planning, the HEAT straddles two fields. The appropriate method may depend on the decision problem itself. Giving the tool user the ability to choose which method they would like to use would be a useful future feature in the tool.

There are several limitations of this study. The biggest perceived challenge to implementing the VSLY in the HEAT is the difficulty users in many countries would face in inputting the age distribution of those affected by an intervention. There is therefore a trade-off between precision and usability in this HIA tool. Potential solutions include (1) using the distribution of age in the general population as a default for the active travel population with the option to manually overwrite or (2) the creation of a bespoke age distribution from user-defined parameters, for example, minimum, maximum, and median age. Although neither of these solutions are perfect, they may provide a compromise between usability and accuracy.

A further challenge exists specifically for the HEAT tool in explaining the VSLY method to stakeholders and users. Transport planners are familiar with the concept of the VSL, but gaining buy-in for the use of the VSLY requires an explanation of how discounted life expectancy is calculated. This is another example of where the adaptation of a widely used tool, already being used by stakeholders to support or inform policy, must be carefully considered even if it is methodologically valid. Over the duration of the HEAT's existence the core team have attempted to achieve balance between complexity and precision on the one hand and usability on the other.[23] However, recent developments in data availability, statistical programming, and web-based user interfaces have made it easier to allow stakeholder engagement in complex models.[24] Therefore, the improvements in the conceptual validity provided by the VSLY method should justify implementation within the global version of HEAT currently under development.

An additional issue for accurate valuation of increased population walking and cycling is that the VSL estimates used (in both the VSL and VSLY methods) are derived from a stated preference study with a median age of 50 years. As VSL has been shown to peak around age 50 years,[11] calculating the VSLY from this figure may result in overestimates. Further research is needed to develop stated preference values that account for the many different factors influencing respondents of different ages.

Conclusion

Our findings suggest that incorporation of duration of life gained into the HEAT is theoretically possible, yields very different results where intervention populations are not representative of overall populations, and is more aligned with guidance from the field of health economics. However, where changes to improve the precision of widely used tools such as the HEAT may also reduce their practical usability, it is important to consider the overall impact on the tool's value to decision-makers and other stakeholders. Thus, it will be important to consider the usability of the modified model in practice in future work.

Author statements

Ethical approval

No ethical approval was required for this simulation study, and all data used are available in the public domain.

Funding

R.S. is jointly funded by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science (108903) and the University of Sheffield. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article.

Competing interests

None.

Authors' contributions

R.S. contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, and writing, reviewing, and editing the article. C.T. and H.S. contributed to supervision and reviewing and editing the article. T.G. contributed to project administration, methodology, and reviewing and editing the article. S.K. contributed to project administration and reviewing and editing the article. E.G. contributed to supervision and reviewing and editing the article. All authors have fulfilled criteria for authorship.

Data availability

Data & Code: https://anonymous.4open.science/r/b1ac653f-7e70-43ab-870c-f3ccc4d63914/

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.puhe.2021.03.016.

References

1. Kahlmeier S, Racioppi F, Cavill N, Rutter H, Oja P. "Health in all policies" in practice: guidance and tools to quantifying the health effects of cycling and walking. *J Phys Activ Health* 2010;**7**:S120–5.
2. Organization WH, others. *Health impact assessment: main concepts and suggested aproach-gothenburg consensus paper*. 2000.
3. Lock K. Health impact assessment. *BMJ* 2000;**320**:1395–8.
4. Mueller N, Rojas-Rueda D, Cole-Hunter T, De Nazelle A, Dons E, Gerike R, et al. Health impact assessment of active transportation: a systematic review. *Prev Med* 2015;**76**:103–14.
5. Waheed F, Ferguson GM, Ollson CA, MacLellan JI, McCallum LC, Cole DC. Health impact assessment of transportation projects, plans and policies: a scoping review. *Environ Impact Assess Rev* 2018;**71**:17–25.
6. Kahlmeier S, Götschi T, Cavill N, Castro Fernandez A, Brand C, Rojas Rueda D, et al. *Health economic assessment tool (HEAT) for walking and for cycling. Methods and user guide on physical activity, air pollution, injuries and carbon impact assessments*. 2017.
7. *Using the health economic assessment tools (HEAT) for walking and cycling: lessons learnt*. World Health Organization; 2017.
8. Brown V, Diomedi BZ, Moodie M, Veerman JL, Carter R. A systematic review of economic analyses of active transport interventions that include physical activity benefits. *Transport Pol* 2016;**45**:190–208.
9. De Nazelle A, Nieuwenhuijsen MJ, Antó JM, Brauer M, Briggs D, Braun-Fahrlander C, et al. Improving health through policies that promote active travel: a review of evidence to support integrated health impact assessment. *Environ Int* 2011;**37**:766–77.
10. Viscusi WK, Aldy JE. The value of a statistical life: a critical review of market estimates throughout the world. *J Risk Uncertain* 2003;**27**:5–76.
11. Aldy JE, Viscusi WK. Adjusting the value of a statistical life for age and cohort effects. *Rev Econ Stat* 2008;**90**:573–81.
12. *Mortality risk valuation in environment, health and transport policies*. OECD; 2012.
13. Woodcock J, Givoni M, Morgan AS. Health impact modelling of active travel visions for england and wales using an integrated transport and health impact modelling tool (ITHIM). *PloS One* 2013;**8**:e51462.
14. Mansfield TJ, MacDonald Gibson J. Health impacts of increased physical activity from changes in transportation infrastructure: quantitative estimates for three communities. *BioMed Res Int* 2015;**2015**.
15. Götschi T, Loh TH. Advancing project-scale health impact modeling for active transportation: a user survey and health impact calculation of 14 US trails. *J Transport Health* 2017;**4**:334–47.
16. Robinson LA, Sullivan R, Shogren JF. Do the benefits of COVID-19 policies exceed the costs? Exploring uncertainties in the age–VSL relationship. *Risk Anal* 2020. https://doi.org/10.1111/risa.13561.
17. Viscusi WK. *Pricing the global health risks of the COVID-19 pandemic. Forthcoming journal of risk and uncertainty*. Vanderbilt Law Research Paper; 2020.
18. Kelly P, Kahlmeier S, Götschi T, Orsini N, Richards J, Roberts N, et al. Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship. *Int J Behav Nutr Phys Activ* 2014;**11**:132.
19. Dicker D, Nguyen G, Abate D, Abate KH, Abay SM, Abbafati C, et al. Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;**392**:1684–735.
20. UN. *World population prospects 2019: highlights (st/esa/ser. A/423)*. 2019.
21. Soetewey A. *Yll: compute expected years of life lost (YLL) and average YLL*. 2018.
22. Nilsson A. *The monetary value of a human life-examining the differences between sectors*. 2014.
23. Götschi T, Kahlmeier S, Castro A, Brand C, Cavill N, Kelly P, et al. Integrated impact assessment of active travel: expanding the scope of the health economic assessment tool (HEAT) for walking and cycling. *Int J Environ Res Publ Health* 2020;**17**:7361.
24. Smith R, Schneider P. Making health economic models shiny: a tutorial. *Welcome Open Res* 2020;**5**:69.

Chapter 4. HEAT - Incorporating a non-linear Dose Response Function.

This chapter contains a publication which reports the results of the second case study using the HEAT in which the net monetary benefit associated with increases in physical activity are compared using a linear dose-response relationship between physical activity and mortality (current HEAT) and a non-linear relationship.

This article [under review] was published open access following the requirement of the Wellcome Trust who financially supported this work. The conditions of the open access publishing allow use of the final published PDF, original submission, or accepted manuscript in this thesis (including in any electronic institutional repository or database). A full set of peer reviewer comments for the entire history of the article can be found online at the link above. All code and data has been provided open source.

https://wellcomeopenresearch.org/articles/7-7/v2

The paper in the chapter was written with 3 co-authors: Chloe Thomas, Hazel Squires and Elizabeth Goyder. Robert Smith was first author and corresponding author and led the Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing (Original Draft Preparation and Review Editing). Chloe Thomas, Hazel Squires and Elizabeth Goyder contributed through supervision and reviewing the manuscript. The paper was conceptualised during my time at the World Health Organisation Collaborating Centre for Physical Activity at the University of Zurich, and I therefore acknowledge Sonja Kahlmeier, Thomas Gostski & Alberto Castro Fernandez for providing details of and access to the HEAT model and data. This work, or variations thereof, does not form a part of any other PhD thesis.

RESEARCH ARTICLE

REVISED A comparison of the World Health Organisation's HEAT model results using a non-linear physical activity dose response function with results from the existing tool [version 2; peer review: 2 approved with reservations]

Robert A. Smith, Chloe Thomas, Hazel Squires, Elizabeth Goyder

School of Health and Related Research, University of Sheffield, Sheffield, South Yorkshire, S1 4DA, UK

v2 **First published:** 07 Jan 2022, **7**:7
https://doi.org/10.12688/wellcomeopenres.17411.1
Latest published: 04 May 2022, **7**:7
https://doi.org/10.12688/wellcomeopenres.17411.2

Abstract

Introduction: The WHO-Europe's Health Economic Assessment Tool is a tool used to estimate the costs and benefits of changes in walking and cycling. Due to data limitations the tool's physical activity module assumes a linear dose response relationship be-tween physical activity and mortality.

Methods: This study estimates baseline population physical activity distributions for 44 countries included in the HEAT. It then compares, for three different scenarios, the results generated by the current method, using a linear dose-response relationship, with results generated using a non-linear dose-response relationship.

Results: The study finds that estimated deaths averted are relatively higher (lower) using the non-linear effect in countries with less (more) active populations. This difference is largest for interventions which affect the activity levels of the least active the most. Since more active populations, e.g. in Eastern Europe, also tend to have lower Value of a Statistical Life estimates the net monetary benefit estimated by the scenarios are much higher in western-Europe than eastern-Europe.

Conclusions: Using a non-linear dose response function results in materially different estimates where populations are particularly inactive or particularly active. Estimating base-line distributions is possible with limited additional data requirements, although the method has yet to be validated. Given the significant role of the physical activity module within the HEAT tool it is likely that in the evaluation of many interventions the monetary benefit estimates will be sensitive to the choice of the physical activity dose response function.

Keywords

Physical Activity, Dose Response Function, HEAT, Walking, Cycling

Open Peer Review

Approval Status ? ?

	1	2
version 2 (revision) 04 May 2022	? view	
version 1 07 Jan 2022	? view	? view

1. **James Woodcock**, University of Cambridge, Cambridge, UK

2. **Marco Hafner**, RAND Europe, Cambridge, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert A. Smith (rasmith3@sheffield.ac.uk)

Author roles: Smith RA: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Thomas C**: Supervision, Writing – Review & Editing; **Squires H**: Supervision, Writing – Review & Editing; **Goyder E**: Funding Acquisition, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: R.S. is joint funded by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science [108903] and the University of Sheffield.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Smith RA, Thomas C, Squires H and Goyder E. **A comparison of the World Health Organisation's HEAT model results using a non-linear physical activity dose response function with results from the existing tool [version 2; peer review: 2 approved with reservations]** Wellcome Open Research 2022, **7**:7 https://doi.org/10.12688/wellcomeopenres.17411.2

First published: 07 Jan 2022, **7**:7 https://doi.org/10.12688/wellcomeopenres.17411.1

REVISED **Amendments from Version 1**

There have been several updates:

1) The Abstract has been tidied to ensure that line-breaks are implemented correctly.

2) There are several adaptations to the text as per responses to reviewers, there has been one additional reference added (see tracked changes).

3) As a result of these changes based on reviewer requests, Figure 3 has been updated in the main body and a new figure and table have been added to the extended data.

Any further responses from the reviewers can be found at the end of the article

Introduction

There is a growing recognition of the importance of considering health in all policies[1–3]. One example of successful integration of health impact in another policy domain is the World Health Organization's Health Economic Assessment Tool (HEAT), which has been widely used, primarily by transport planners, to estimate the health benefits associated with increased walking and cycling[4]. The success of the HEAT is in part due to its simplicity, requiring relatively few user inputs compared to other health economic models[5].

However, a limitation of the HEAT is that despite broad consensus that the relationship between physical activity and all-cause mortality is non-linear, such that the greatest health benefits from an extra unit of physical activity accrue in those who are least active[6–8], the HEAT assumes a linear relationship between physical activity and mortality. The HEAT methods and user guide states that "a linear relationship was chosen to avoid additional data requirements on baseline activity levels (which would be needed using a non-linear dose–response function)"[4]. There is however a recognition that improvements in data availability could allow for a non-linear relationship to be used in the future. The same report states that "An approach based on a non-linear relationship could be adopted as part of future updates of HEAT, when suitable data on the baseline level of physical activity in different populations are available to provide default values for HEAT" (p.9).

This study uses a method developed by Hafner *et al.*[9] to estimate the distribution of physical activity in 44 countries in the WHO European Region for which the HEAT applies. It then compares, for three hypothetical scenarios, the number of deaths averted and the monetary benefit when assuming a linear relationship, as done by the current HEAT model, and a non-linear relationship between physical activity and all-cause mortality. Although previous analysis has shown the importance of estimating changes in the distribution of physical activity, rather than categorizing activity levels[10] this is the first time that the effect of the shape of the dose response relationship has been analysed within a single health economic model, with all other structural assumptions held constant. Woodcock *et al.* (2013)[11] estimated the difference in the number of deaths averted between the ITHIM and HEAT tools when modelling all-cause mortality, and when modelling several diseases individually. Since the ITHIM model uses a non-linear power transformation, the difference between the ITHIM and HEAT does in part reflect differences associated with the dose-response function. However, there are other differences between the ITHIM and HEAT which make it impossible to isolate the effect of the shape of the dose-response relationship for physical activity on model outcomes. This study aims to isolate this effect, to investigate how sensitive the HEAT model is to the assumed dose response relationship.

Material and methods

Data and measures

This study uses data on the prevalence of insufficient physical activity in 44 HEAT countries from a publication by 12, the self-reported non-occupational (leisure time and commuting) physical activity levels of a representative sample of the English population from the Health Survey for England 2015[13], country specific mortality rates for those aged 20–74 from the European Mortality Database[14] and value of a statistical life estimates from a systematic review[15]. It uses the linear dose-response relationship between physical activity and mortality from 7 as described in the HEAT methodology paper[4], and a non-linear dose-response relationship as described in 16. A summary of data including sources can be found in Table 1.

Analysis

We estimate the number of deaths averted per 100,000 and the net monetary benefit using both the non-linear dose-response method and the linear dose-response currently used by HEAT for 44 European countries in three scenarios:

1. Scenario 1: An extra 10 minutes of daily walking for every person in the population.
2. Scenario 2: Every adult meets WHO Guidelines. Every adult in the country who doesn't already meet WHO guidelines of 600 MET-mins per week (equivalent to around 150 minutes of brisk walking per week) increase their activity to that level. Those meeting guidelines are unchanged.
3. Scenario 3: A 10% increase in physical activity levels of the population aged 20–74, such that those who are the most active have the largest absolute activity increase, and those who are least active have the smallest absolute activity increase.

This analysis is not an attempt to estimate the probability, feasibility or costs of achieving the scenarios. For each scenario we assume that the outcome is achieved, and we estimate the benefits in terms of deaths averted per 100,000 and monetize these benefits using the VSL.

The current HEAT method using a linear dose response relationship

The current HEAT method requires the user to input pre-intervention and post-intervention physical activity levels, in terms

Table 1. Variable names, description and source of data used in analysis.

Variable	Description	Source
From the HEAT methodology		
MR	Country specific mortality rates (for ages 20–74)	European Mortality Database (2017)[14]
RR_lit	Relative risk in literature	7; 4
mins_ref	Reference physical activity duration	7; 4
RR_min	Minimum relative risk (max effect)	7; 4
VSL	Value of a Statistical Life for each country	15
Other sources		
piap	% of population inactive	12 Appendix 5
mets	Distribution of met mins in English population	HSE 2015[13]
t	Log-linear dose response function power (t)	16

of minutes of walking and cycling[4]. It estimates the relative risk associated with each activity level using Equation 1 below.

$$R = max\{1-(1-RR_{lit})\times\frac{Mins_{local}}{Mins_{ref}}, RR_{min}\} \quad (1)$$

For a walking intervention the relative risk RR_{lit} is 0.89, the reference minutes of activity from the literature $Mins_{ref}$ is 168mins per week and the risk reduction cap RR_{min} is 0.7, such that every additional 10 minutes of weekly walking ($Mins_{local}$ = 10) reduces relative risk by 0.65 percentage points, to a limit of 30 percentage points. Number of deaths averted *DA* is then calculated by multiplying the absolute difference in relative risk between intervention and comparator (RR_i–RR_c) by the country specific mortality rate of the population aged 20–74 MR_c and the population affected, *pop* (Equation 2 below). This is then monetized in terms of monetary benefit (MB) in Equation 3 by multiplying the number of deaths averted by the country specific value of a statistical life VSL_c:

$$DA = (RR_i - RR_c)\times MR_c \times pop \quad (2)$$

$$MB = DA \times VSL_c \quad (3)$$

The adapted method using a non-linear dose response relationship

The non-linear dose response method requires a baseline distribution of physical activity. We use weekly metabolic equivalent of task minutes (MET-minutes) from moderate and vigorous physical activity to summarize an individual's physical activity level in one number[17]. A distribution of weekly MET-mins for each country was imputed using a method from Hafner *et al.* (2019)[9]. This method combines estimates of prevalence of physical inactivity for each of the 44 countries with the distribution of physical activity in a generic distribution (we use the distribution derived from the Health Survey for England). Each percentile, *n*, of physical activity in the target country, *c*, distribution is calculated separately using the equation below.

$$p_c^n = p_c^{n-1} + (p_g^n - p_g^{n-1})\frac{x_c}{x_g} \quad (4)$$

The weekly MET-mins, *p*, for each country *c*, at each percentile, n, is based on the prevalence of sufficient physical activity, x_c, in the country, *c*, compared to the prevalence of sufficient activity in the generic distribution, x_g. The values for each percentile then form the estimated physical activity distribution for each country. More detail on this method, as well as comparisons of country distributions, can be found in Hafner *et al.* (Appendix C)[9]. The estimates derived from these equations, along with a density plot for 6 countries included in the analysis, are available in the extended data.

The population relative risk is calculated as the simple arithmetic mean of relative risk for each percentile of the physical activity distribution, as shown in Equation 5 below. For each percentile relative risk is estimated using a log-linear relationship, calculated using the relative risk from the literature (RR_{lit} = 0.89), percentile MET-mins ($mets_p$), reference MET-mins ($mets_{ref}$) which is simply 4 (METs associated with moderate physical activity) × $mins_{ref}$ (from Kelly *et al.*, 2014[7]), and a power transformation *t*. The power transformation is 0.375 in the main analysis, and varied from 0.25 to 0.75 in sensitivity analysis, as recommended in Woodcock *et al.*[16].

$$RR = \frac{\sum_{(p=1)}^{(N=100)} RR_{lit}^{(\frac{mets_p}{mets_{ref}})^t}}{100} \quad (5)$$

Figure 3 shows the dose response relationship between physical activity and all cause mortality risk for the linear model and the non-linear models with different values of t is shown in the extended data.

Once relative risk is calculated, the deaths averted and monetary benefit are calculated using Equation 2 and Equation 3.

Comparison

For each of the 44 countries included in the analysis, for each of the three scenarios, and for each of the four dose response functions, we calculated two metrics:

- the number of deaths averted per 100,000 persons aged 20–74.

- the monetary benefit associated with mortality reduction, using the HEAT VSL estimates for each country[15].

A comparison of the number of deaths averted under different modelling methods are displayed using simple scatter plots with a 45-degree line of equality, and monetary benefit estimates are shown, in Euros, on choropleth maps of Europe.

Results

The estimated distributions of physical activity for each of the 44 countries in the analysis are provided in the extended data, and can also be found on this GitHub repository: https://github.com/RobertASmith/HEAT_DRF/blob/master/output/country_mets_dist.csv.

A comparison of the number of premature annual deaths averted per 100,000 people using the two different methods in each of the three scenarios for the 44 WHO European Region countries is shown in Figure 1 below. The estimates derived using the linear dose response method are shown on the x-axis and the non-linear dose response on the y axis. A 45-degree line of equality is plotted to aid comparison. The country points are labelled with ISO3 codes and shaded from black for low insufficient physical activity prevalence (IPAP) to blue for those with a high IPAP.

The figure shows that for the first scenario, an additional 10 minutes of daily walking, countries with particularly inactive (active) populations tend to have higher (lower) estimated deaths averted using the non-linear function compared to the linear function.

In the second scenario all individuals with activity levels below WHO physical activity guidelines of 600 MET-mins per week increase activity to meet guidelines. Here, the non-linear function results in higher deaths averted than the linear function in most countries, except for some with especially low prevalence of insufficient physical activity (e.g. Moldova and Belarus).

In the third scenario, in which all individuals increase their physical activity level by 10%, estimates derived using a non-linear function are much lower than using a linear function for all countries, regardless of the prevalence of insufficient physical activity. This is because those with low physical activity levels, who would benefit the most from increased physical activity according to a non-linear model, have low increases in MET-mins, while those who are highly active have high absolute increases in MET-mins but benefit little in terms of premature mortality reduction when using a non-linear model.

In order to allow for trade-offs in decision making between health and non-health outcomes, the HEAT tool monetises the deaths averted using the Value of a Statistical Life (VSL)[18], giving an estimate in terms of monetary benefit. Figure 2 below shows the monetary benefit associated with Scenario 1, using a log-linear dose response function with a power transformation of 0.375. The monetary benefits tend to be higher in countries with higher insufficient physical activity prevalence and higher VSL (e.g. Ireland, the UK and Luxemburg) and markedly lower in countries with lower VSL and/or lower physical inactivity prevalence such as Ukraine and Moldova, this results in marked differences between the West and East Europe.

Discussion

Increasing population physical activity is likely to yield large benefits in health, wellbeing & productivity worldwide[9]. However, trade-offs often exist between increasing population physical activity and achieving other health and non-health outcomes. It is therefore important to have a robust method to consider whether interventions that improve activity levels provide good value for money. The HEAT is an example of a tool, often used by transport planners, which allows users to estimate, and monetize, the benefits of increased walking and cycling[3]. In general, the estimates derived from the physical activity module of the tool have been shown to contribute the most to total monetary benefit (Mueller *et al.*, 2015).

We describe an adaption to the current HEAT physical activity module which applies a non-linear dose response relationship between physical activity and mortality risk to estimated country specific baseline distributions of physical activity. The method is more sensitive to interventions which increase the activity levels of the least active, and less sensitive to interventions which increase the activity levels of the most active. This means that similar scenarios may yield less health benefit in more active countries. As noted in our previous work[5], since countries with higher GDP tend to have a higher Value of a Statistical Life[15] and higher prevalence of insufficient physical activity[12], the estimated net monetary benefit tends to be higher in western Europe than eastern Europe.

There are numerous limitations of this analysis. Firstly, the method used to estimate the baseline distributions of physical activity in each of the HEAT countries (from 9) assumes that the shape of the physical activity distribution is relatively similar in every country. Comparing the distributions estimated by this method, and provided in the extended data, with more detailed datasets would help to validate the estimates of population physical activity distributions. It is likely that the method is reliable for similar countries (e.g. the UK and Germany) but may not be reliable where culture differs

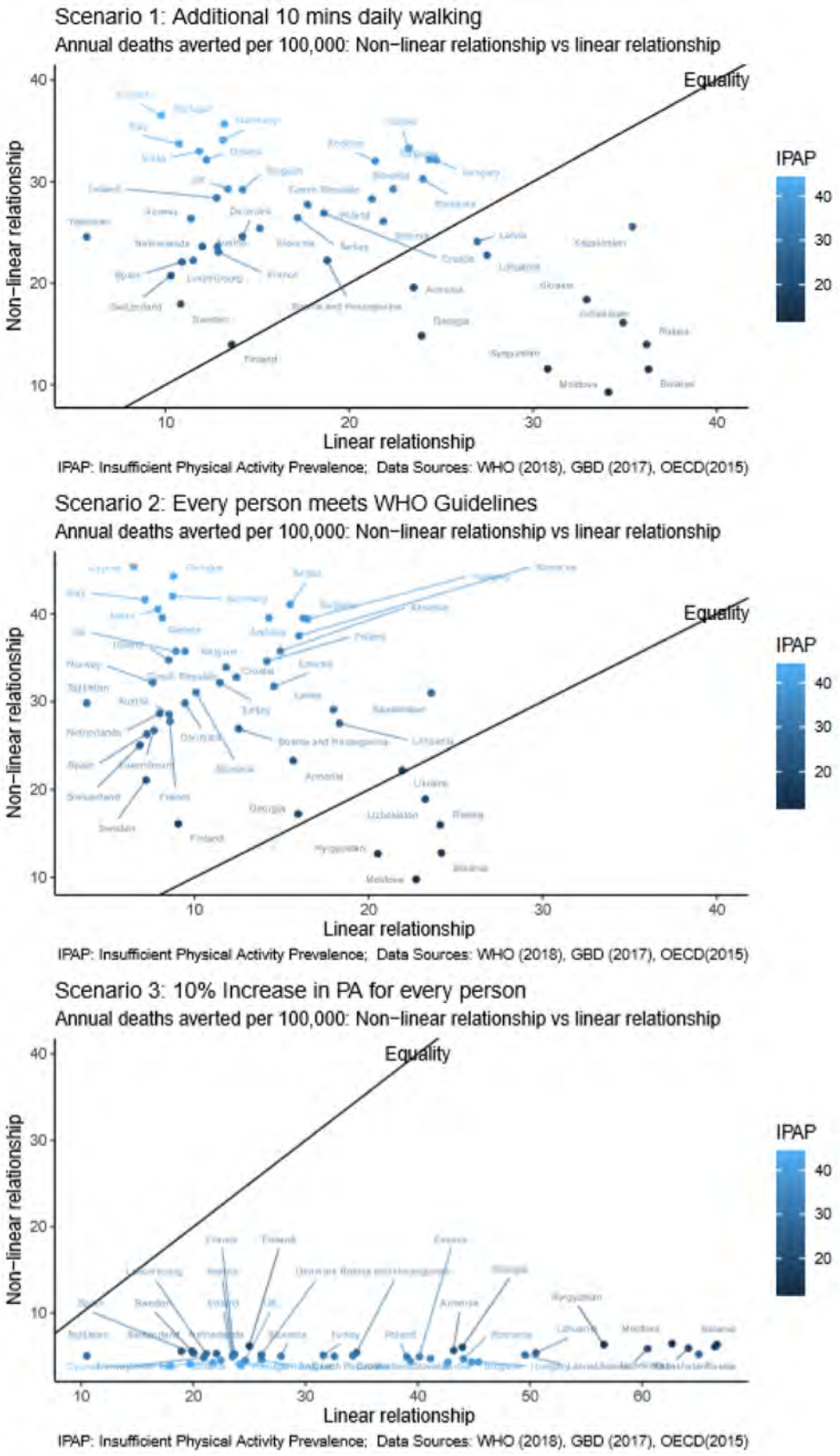


Figure 1. Deaths averted per 100,000 for three scenarios using the non-linear and the current (linear) relationship.

Figure 2. **Annual Monetary Benefit of an additional 10 minutes daily walking for 44 European Countries, in 2016 USD.**

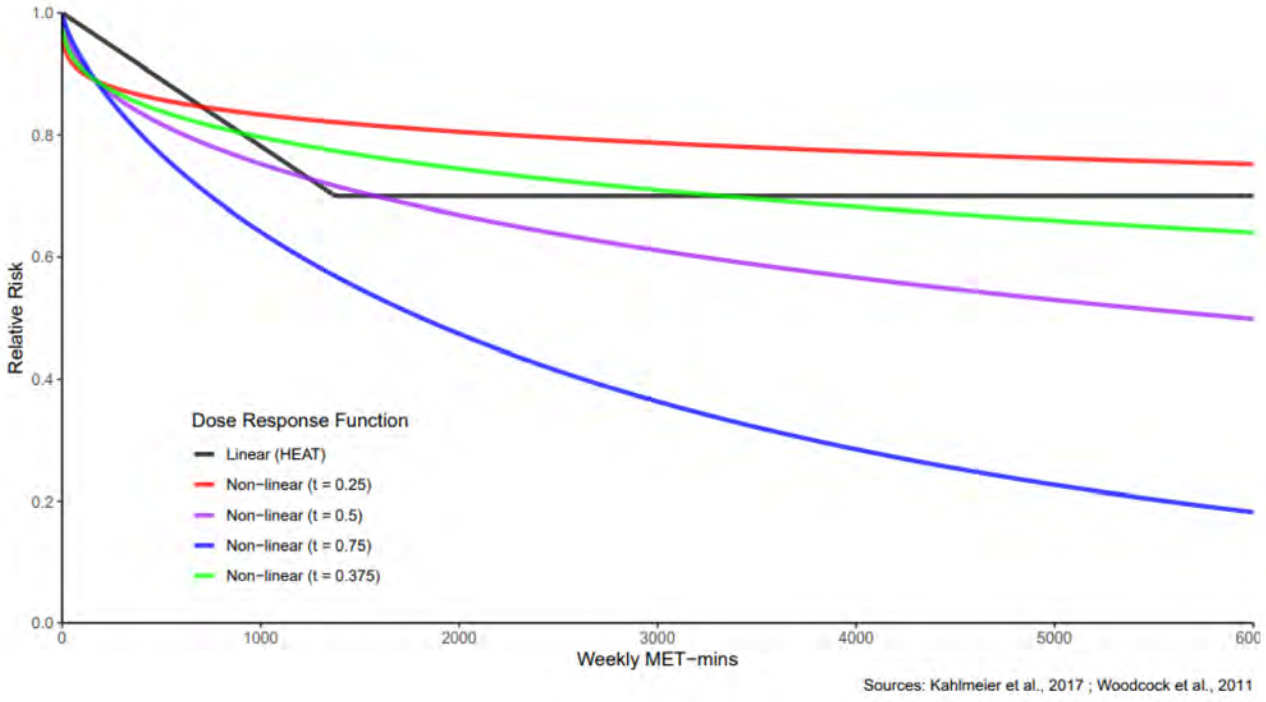


Figure 3. **Relative risk using linear & non-linear dose response functions with different power transformations:**

(e.g. the UK and Chad). However, it is unlikely that this would affect the main finding of this study, since large differences in the linear and non-linear functions exist when using the UK distribution which is based upon survey data. It is also worth noting that in this study, as in many other studies relying upon secondary data, the assumption is made implicitly that the same survey methods for physical activity are utilised in the estimation of the dose-response function, and for the purposes of calculating relative risk. Any differences in the survey methods will generate a bias in the estimation of relative risk.

We also note that the comparison between the linear dose response function currently used by the HEAT, and a non-linear function based on Woodcock *et al.* 2011 is a false dichotomy. It is likely that non-parametric regression techniques, such as spline regression will yield a dose response relationship that is more appropriate, avoiding implausibly large benefits for particularly inactive individuals, which is apparent at low levels of Weekly MET-mins in Figure 3. However, the authors are not aware of any such studies directly relating to population mortality to date, although it is likely new evidence will emerge.

A further limitation of this study is that we do not consider the usability of the tool, only show that a more conceptually valid method is possible. Since the tool is designed to be used by users with little to no public health, epidemiology, statistics and programming ability it is also important that the methods behind the tool are easy to explain, and the tool is simple to use. Increased complexity, in terms of more, or more detailed inputs, and a more difficult to explain model structure may make the tool less 'use-able', used here as a loose term which encompasses both technical feasibility of use and user understanding & confidence, and therefore less valuable. Further work to determine whether stakeholders understand the use of a non-linear dose response relationship on baseline and intervention distributions, and whether users can obtain intervention group physical activity distributions, will likely be a determining factor as to the feasibility of adapting the HEAT tool. Nevertheless, this paper demonstrates that the two approaches do result in substantial differences at the population level, and therefore where possible the non-linear dose response function should be used by researchers.

The trade-off between the 'usability' and 'accuracy' of health impact assessment tools (and public health economic models more generally) is one that needs further attention in the academic literature. Models and tools tend to be either high accuracy but low usability - for example models created in high level programming languages with high computational demands and long runtimes - or low accuracy but high usability - including the HEAT physical activity modules. Understanding how to utilize new tools from data-science to make models which are very accurate and usable would be a useful avenue of future research. Likewise, understanding how to incrementally improve the accuracy of highly usable models (like HEAT) without compromising usability would be a valuable endeavor.

Conclusions

We show that for the WHO European Region countries included in the HEAT tool, the estimates of deaths averted, and therefore monetary benefit, differs substantially depending on the dose response function used. The nonlinear dose response function results in greater estimated benefits, relative to the linear dose response function, where increased physical activity accrues to those who are relatively inactive. It therefore results in greater benefits in countries with higher prevalence of physical inactivity, or interventions which are targeted toward the least active. Developing tools which are both usable, in terms of data requirements and ease of explanation to users, and highly accurate is an important avenue for future research in health impact assessment and public health economics more widely.

Data availability

Zenodo: A comparison of the World Health Organisation's HEAT model results using a non-linear physical activity dose response function with results from the existing tool. https://doi.org/10.5281/zenodo.6505091.

This project contains the following underlying data:

Data file 1. (Density Plot).

Data file 2. (Country physical activity distributions used in the analysis).

Author contributions

Robert Smith: Conceptualization, Methodology, Visualization, Investigation, Writing-Original draft preparation.

Chloe Thomas: Supervision, Writing- Reviewing and Editing.

Hazel Squires: Supervision, Writing- Reviewing and Editing.

Elizabeth Goyder: Supervision, Writing- Reviewing and Editing.

Acknowledgements

The authors would like to thank Martin Stepanek & Marco Hafner for detailed explanations explaining their method to estimate country physical activity distributions. We would also like to thank Sonja Kahlmeier, Thomas Gostski & Alberto Castro Fernandez for providing details of and access to the HEAT model and data.

References

1. World Health Organization *et al.*: **Health in all policies: Helsinki statement. framework for country action.** 2014.
Reference Source

2. Koivusalo M: **The state of health in all policies (hiap) in the european union: potential and pitfalls.** *J Epidemiol Community Health.* 2010; **64**(6): 500–503.
PubMed Abstract | Publisher Full Text

3. Kahlmeier S, Racioppi F, Cavill N, *et al.*: **"Health in all policies" in practice: guidance and tools to quantifying the health effects of cycling and walking.** *J Phys Act Health.* 2010; **7 Suppl 1**(s1): S120–S125.
PubMed Abstract | Publisher Full Text

4. Kahlmeier S, Götschi T, Cavill N, *et al.*: **Health economic assessment tool (heat) for walking and for cycling. methods and user guide on physical activity, air pollution, injuries and carbon impact assessments.** 2017.
Reference Source

5. Smith R, Thomas C, Squires H, *et al.*: **The price of precision: trade-offs between usability and validity in the world health organization health economic assessment tool for walking and cycling.** *Public Health.* 2021; **194**: 263–269.
PubMed Abstract | Publisher Full Text

6. Arem H, Moore SC, Patel A, *et al.*: **Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship.** *JAMA Intern Med.* 2015; **175**(6): 959–967.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Kelly P, Kahlmeier S, Götschi T, *et al.*: **Systematic review and meta-analysis of reduction in all-cause mortality from walking and cycling and shape of dose response relationship.** *Int J Behav Nutr Phys Act.* 2014; **11**(1): 132.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Ekelund U, Tarp J, Steene-Johannessen J, *et al.*: **Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis.** *BMJ.* 2019; **366**: l4570.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Hafner M, Yerushalmi E, Phillips WD, *et al.*: **The economic benefits of a more physically active population: an international analysis.** 2019.
Publisher Full Text

10. Minton J, Dimairo M, Everson-Hock E, *et al.*: **Exploring the relationship between baseline physical activity levels and mortality reduction associated with increases in physical activity: a modelling study.** *BMJ Open.* 2013; **3**(10): e003509.
PubMed Abstract | Publisher Full Text | Free Full Text

11. Woodcock J, Givoni M, Morgan AS, *et al.*: **Health impact modelling of active travel visions for England and Wales using an Integrated Transport and Health Impact Modelling Tool (ITHIM).** *PLoS One.* 2013; **8**(1): e51462.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Guthold R, Stevens GA, Riley LM, *et al.*: **Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants.** *Lancet Glob Health.* 2018; **6**(10): e1077–e1086.
PubMed Abstract | Publisher Full Text

13. NHS Digital: **Health Survey for England.** London, UK. 2016.
Reference Source

14. MDB (European detailed mortality database): **Copenhagen: WHO Regional Office for Europe.** 2017; accessed 18 October 2017.

15. **Mortality Risk Valuation in Environment OECD: Health and Transport Policies.** OECD Publishing, Paris. 2012.
Publisher Full Text

16. Woodcock J, Franco OH, Orsini N, *et al.*: **Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies.** *Int J Epidemiol.* 2011; **40**(1): 121–138.
PubMed Abstract | Publisher Full Text

17. Ainsworth BE, Haskell WL, Whitt MC, *et al.*: **Compendium of physical activities: an update of activity codes and met intensities.** *Med Sci Sports Exerc.* 2000; **32**(9 Suppl): S498–S504.
PubMed Abstract | Publisher Full Text

18. Viscusi WK, Aldy JE: **The value of a statistical life: a critical review of market estimates throughout the world.** *J Risk Uncertainty.* 2003; **27**(1): 5–76.
Publisher Full Text

Chapter 5. HEAT Model - Discussion

In the chapters above, I discussed two ways in which the HEAT tool, which has been shown to be particularly useful, can be adapted to improve its external validity.

According to the first paper, which compared different approaches for valuing deaths averted in the model, the use of the VSLY method allows for a more nuanced analysis of the benefits of increased physical activity. While estimating the number of deaths averted by a policy is useful to decision makers, it does not consider the duration of lives saved. The VSLY method, on the other hand, accounts for both the number of lives saved and the length of time that each life is extended, providing a more complete picture of the overall benefit of an intervention. This is particularly important when interventions which affect different populations (older vs younger persons) are competing for scarce resources. The paper, published in *Public Health,* argued that by assigning the same value to mortality risk reduction to all groups, regardless of age, we are implicitly assigning a higher value on each year of life for older people, something that may be perceived as unfair by those who think that, all else equal, an extra (statistical) year of life should be valued equally regardless of the age of the recipient.

As noted in the paper "*The price of precision*," making changes to improve the precision of tools like the HEAT "*may also reduce their practical usability*" (p.268). In health economic modelling, there is sometimes a trade-off between model usability and external validity. More externally valid models may require more complex data and be harder to use and interpret, while simpler models may be easier to use but less externally valid. In the case of incorporating a life-years approach, this may require a more detailed understanding of the age distribution of those affected by the intervention, which may not be feasible for all users of the tool. It is important to carefully consider this trade-off between usability and external validity in order to ensure that the model meets the needs of the intended audience and serves its intended purpose.

In the second paper I discussed the use of a non-linear dose-response function to replace the linear dose response function currently used in the HEAT. We know that the benefits to a unit of increased physical activity are larger for those who are less active (Geidl et al. 2020; Grandes et al. 2023). The proposed changes allow this to be more accurately captured by the HEAT tool model: "*The nonlinear dose response function results in greater estimated benefits, relative to the linear dose response function, where increased physical activity accrues to those who are relatively inactive*". This has important implications for policy, since interventions that focus on encouraging people who are currently inactive to become more active will, all else being equal, be the most cost-effective use of resources and may reduce inequalities.

Despite the potential benefits of using a nonlinear dose-response function in the HEAT tool, there are still significant challenges to implementing this approach. One challenge is determining the shape of the nonlinear function, which can be more complex than determining the slope of a linear function

(Kahlmeier et al. 2017). Another challenge is accurately estimating the distribution of baseline physical activity in the population affected by the intervention. This can be difficult to do, as survey methods for collecting this information can be expensive and people may overestimate their physical activity levels in surveys. These challenges must be addressed in order to effectively incorporate a nonlinear dose-response relationship into the HEAT tool.

The common theme

A lot of time and resource is devoted to developing a model structure and parameter inputs which capture most or all of the important characteristics of a system or process that impact on outcomes which may influence decisions (Squires et al. 2016b). We refer to the extent to which this is achieved as ”external-validity”, which can be evaluated in several ways, including comparing the predictions of the model to real-world data, assessing the internal consistency of the model, and comparing the results of the model to the results of other models that have been developed to study similar questions. No model is perfectly accurate, and all models are subject to limitations and assumptions .

However, ultimately, the goal of a health economic model is typically to provide useful and relevant information to inform decision-making (Huserau et al. 2013). As a result, there is often a trade-off between usability and external validity. More externally valid models may require more complex and detailed input data, and may be more difficult to use and interpret, or to adapt to different questions (Hoffmann et al. 2002). On the other hand, simpler models may be easier to communicate, and therefore may gain credibility with decision makers who find them easier to engage with (to ‘use’) but may be less externally valid. This trade-off can be challenging because models that are hard to communicate or difficult for decision makers to interact with may not be practical or useful for decision-making, while models that are too simple or inaccurate may not provide reliable or relevant information.

It is important for modellers to carefully consider this trade-off and choose the model that is most appropriate for their intended audience and research question. Caro et al. (2012) suggest that “finding the balance between simplicity of modelling and avoidance of oversimplification [...] is perhaps the most important skill that a modeller can learn if a model is to truly fulfil its potential as a communication tool” (p.675). We extend this more widely to usability vs external-validity, whereby one must consider both and may choose to prioritise one over the other depending on the specific needs and goals of the decision maker - in a similar fashion to a productive possibility frontier in classical economic theory (Pareto, 1906).

There are several factors that could yield improvements in external-validity or usability without commensurate reductions in the other. Some of these factors include:

- An improvement in the algorithms or methods used to build the model which simultaneously better reflect reality and are more intuitive and easier to explain. For example, to some extent individual level simulations which follow patients through a journey can be easier to explain to decision makers (Brennan et al. 2006) .

- An increase in the computing power available to run the model. This could allow the model to handle larger datasets or more complex calculations, improving its external validity without sacrificing usability. Likewise, it could reduce the run-time of an existing model.
- A reduction in the resources required to build or maintain a model. This could allow for the development of more externally valid models that are still affordable and practical for decision-makers to use. This may be driven by a decrease in the hourly cost of labour or an increase in the efficiency of labour.
- Improvements in technology which improve the user-friendliness of the model's interface or documentation or increase in the level of training or support available to users of the model, or a given resource. These improvements could help users to better understand and use the model, improving its usability without sacrificing external validity.

In chapters 7 and 8 we will discuss the disruptive role of new technologies, in particular methods from data-science and computing, in changing this trade-off, and in shifting the production possibility frontier outwards. The focus will be in the last two points above: increasing the level of *usability* of a model and reducing the resources required to develop and maintain models.

Part 3 - Developing a new model: a case study in modelling the long run cost effectiveness effect of school based physical activity interventions

Chapter 6. Physical Activity in Children Economic Model (PACEM)

Abstract

This chapter develops and tests new methods to increase the external validity of health economic evaluations of physical activity interventions in children and young adults. It uses a school based physical activity intervention as a case study, describing a new microsimulation model which incorporates a non-linear dose response relationship between physical activity and multiple health conditions simultaneously. As well as comparing the findings to previously published evaluations of similar interventions, it investigates the sensitivity of the model to different methods of estimating health-related quality of life and to structural assumptions about decay of intervention effect and habit formation. The model results are very sensitive to habit formation and assumptions around the duration over which the health benefits of physical activity last. This was anticipated since much of the health burden of physical activity related diseases occurs in older adults. The chapter concludes by discussing the implications of the findings for decision modellers attempting to estimate the cost-effectiveness of interventions to increase physical activity, much of which applies to other public health interventions. In particular, since other health economic evaluations attempt to model costs and benefits via similar pathways, it is likely that they are also sensitive to the same modelling choices.

Introduction

As discussed in Chapter 1, physical activity is crucial for maintaining overall health and well-being. Numerous studies have shown that increasing physical activity can significantly reduce the risk of various health conditions and decrease all-cause mortality (Warburton et al. 2017). However, despite

the clear benefits of physical activity, many children worldwide do not engage in enough physical activity and fail to meet recommended guidelines (Guthold et al. 2020). Furthermore, physical activity levels tend to decrease with age, although research indicates that physical activity habits tend to persist throughout the lifespan, with individuals who were most active during childhood and young adulthood more likely to be active in middle and old age (Lounassalo et al. 2019; van Sluijs et al. 2021).

Interventions which are effective at increasing physical activity in childhood and adolescence and result in long-term habit formation are likely to be a good investment. However, while the costs of these interventions can be easily observed in the short term, the benefits occur over a longer period of time and are therefore difficult to quantify directly. As discussed in Chapter 2, to estimate the long-term cost-effectiveness of interventions, health economic models are commonly used. Such models are constructed with certain simplifying assumptions, including the linear decay of an intervention's effect, the proportion of the population whose habits are sustained, the duration of physical activity's effect on health outcomes, and the simplification of physical activity into an ordinal variable with "states" that people can transition between (Candio et al. 2012).

It is important to understand how the methodological assumptions used in these models affect their outcomes, to make informed policy and resource allocation decisions. Previous studies have performed sensitivity analysis on their results, but typically these are often not the primary focus of study (Gc et al., 2019; Frew et al., 2014). This chapter investigates the impact of some of these assumptions on the cost-effectiveness of interventions aimed at increasing physical activity in children and young adults. A school based physical activity intervention is used as a case study, the effect size of which is derived from a meta-analysis of school-based interventions conducted by Mears & Jago (2016). This intervention is the same as one analysed in Gc et al. (2019), which enables some comparison of findings between modelling studies.

To test the effect of structural and parametric assumptions on the cost-effectiveness estimates, a new model is developed. The model is distinct from the model developed by Gc et al., but uses the same parameter values where possible, to allow for a comparison of the model results. It deviates from Gc et al. by using a microsimulation model (rather than a Markov model) to incorporate a non-linear dose response relationship between physical activity and health outcomes, captures multiple health conditions simultaneously, estimates incremental discounted costs and Health Related Quality of Life (utilities) and incorporates individual level subgroup effects.

Sensitivity analysis is conducted to investigate the effect of assumptions about decay of intervention effect and habit formation in the intervention population. It investigates the effect of different methods used to estimate health-related quality of life, assessing the sensitivity of the model to multiplicative and minimum methods for health state quality of life aggregation previously discussed in Ara & Brazier (2010). It also investigates the sensitivity of the model to both intervention cost and effect size through economically justifiable costing analysis. Some of the sensitivity analyses are similar to those conducted

by Gc et al. and other models previously, allowing for some inference as to whether some parameters and structural assumptions may be particularly important for physical activity models.

Methods

A probabilistic individual-level microsimulation model was developed to estimate the costs and benefits of interventions which aim to increase physical activity levels in children and adolescents relative to a *No intervention* scenario. The scope (the population, intervention, comparator and outcomes) is described first, after which the structure is outlined in detail. The methods used to describe the probabilistic analysis and the one-way sensitivity analysis are then outlined. Finally, access to open-source code, data and validation methods are described.

The model operates in fixed one-year cycles, with each individual's physical activity trajectory impacting their risk of developing cardiovascular, metabolic and oncological diseases, and therefore directly and indirectly, their risk of dying in any given cycle. The model does not take into account interactions or competition for resources among individuals. It calculates individual health utilities and costs from the perspective of the National Health Service (NHS) and Personal Social Services (PSS) from age 11 to 80 and aggregates them for the entire cohort. It is probabilistic, with 5,000 iterations run, with each sampling from underlying parameter distributions. The mean discounted costs and quality-adjusted life years (QALYs) are reported as primary outcomes with information on incidence and prevalence of the five diseases also provided.

The process of developing the model followed the framework described in Squires et al. (2016b). The framework is made up of four phases: Aligning the framework with the decision-making process, identifying relevant stakeholders, understanding the problem, and developing and justifying the model structure. This *'problem-oriented conceptual model'* (conceptual model) (Tappenden, 2012) that resulted from this process is shown in Appendix A4. This was a result of incorporating the outcome of the reviews described in Chapter 1 and 2, with input from an expert advisory group consisting of clinicians and experts in physical activity modelling and health economics from 2017 to 2019. The scope of the model was subsequently reduced due to limitations in data and an unexpected reduction in capacity. As a result, some of the outcomes (e.g. mental health outcomes and academic attainment), mediating variables (e.g. BMI) and subgroup analysis (IMD quintile) were removed when developing the '*design-orientated conceptual model'* (Tappenden, 2012), which is described in more detail below. This is unfortunate given that many of these were identified in Chapter 1 as being important determinants, and in Chapter 2 as being neglected by previous studies.

The methods were developed to correspond to the CHEERs checklist, a set of guidelines for reporting economic evaluations of healthcare interventions that covers key components such as model structure,

data inputs, assumptions, and transparency (Husereau et al. 2022), a completed version of which can be found in Appendix A2.

The model scope

The model population is a cohort of 10,000 children aged 11, sampled to be representative of the general population in sex, socioeconomic deprivation, and physical activity level. The cohort size of 10,000 was specifically chosen to ensure enough statistical power to capture relatively rare events, such as the occurrence of cancers in younger persons, and to provide reliable estimates of their frequency and characteristics in the population. The model's time horizon is 69 years, from age 11 to age 80. The intervention is a hypothetical after-school intervention with an effect size based on a meta-analysis of similar interventions published by Mears & Jago (2016) and a cost obtained from Gc et al. (2019). The comparator is the current status quo, which assumes that the plethora of interventions and initiatives currently in place result in a decay of PA in adolescence similar to that observed in epidemiological studies (Dumith et al. 2011a). The perspective taken is a National Health Service (NHS) and Personal Social Services (PSS) perspective. The primary model outcomes are incremental discounted Quality Adjusted Life Years, and incremental discounted healthcare costs. The evaluation is therefore a cost-utility analysis.

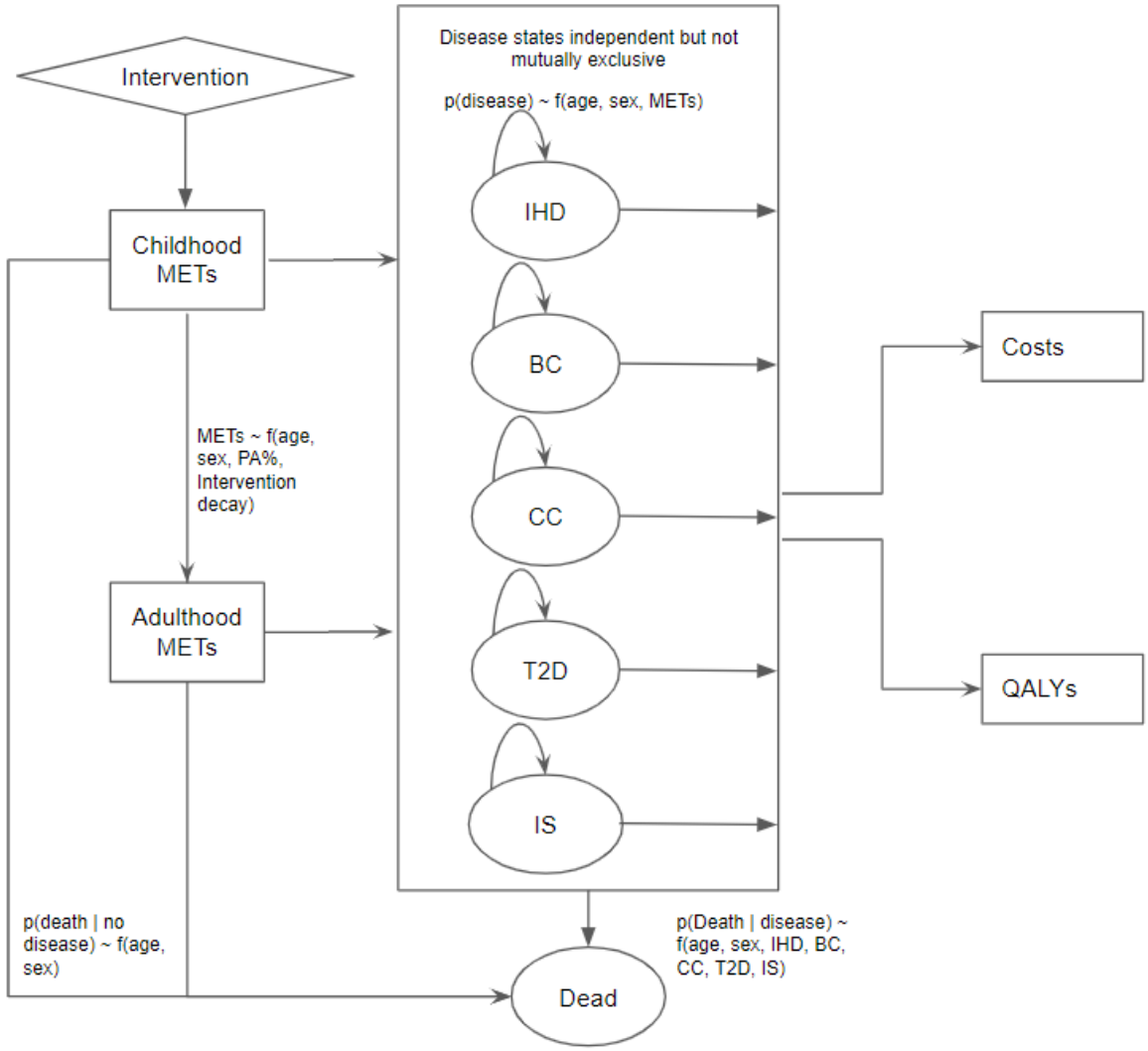
Five diseases were included in the model: Ischemic Heart Disease (IHD), Breast Cancer (BC), Colorectal Cancer (CC), Type 2 Diabetes (T2D) and Stroke (IS). These diseases were chosen because (1) they are the conditions most commonly included in health economic models of physical activity, as discussed in Chapter 2, and (2) epidemiological studies provide good quality evidence on the incidence, prevalence and non-linear dose response relationship with physical activity (Kyu et al. 2016). The original intention was to incorporate mental health outcomes (anxiety and depression) into the model, a key limitation of other models identified in Chapter 2, as well as musculoskeletal injuries. However, the evidence-base is not as strong for these intermediate outcomes as for the five 'core' conditions included, and the scope of this chapter was reduced due to time pressures, and therefore they were not included. A further discussion of this limitation is included in the limitations section.

The model structure

The model is a closed system, meaning that the cohort of children is followed over time, and no new individuals are added. The model consists of a physical activity module that creates trajectories of physical activity over the lifecourse, an epidemiological module that estimates disease and survival status for every individual in each cycle, and an economic module that calculates health state utilities and costs accrued by each individual over the lifecourse. A diagram of the model structure can be seen in Figure 2 below. Each child's initial physical activity level (MET-mins per week), sex and IMD quintile are sampled jointly to ensure the population is representative of the general population at age 11. Physical activity level (METs) in each cycle is dependent on age, sex, no-intervention scenario PA

distribution, and intervention effect. Age, sex and METs then affect (independently) the risk of developing Ischemic Heart Disease (IHD), Breast Cancer (BC), Colorectal Cancer (CC), Type 2 Diabetes (T2D) and Stroke (IS). Individuals can die at any age, with their mortality risk being a function of age, sex and disease status. The duration of each disease is fixed, and physical activity only influences healthcare costs and HRQoL via disease status.

Figure 2 Diagram of PACEM model structure.



Physical activity module

Initial population physical activity

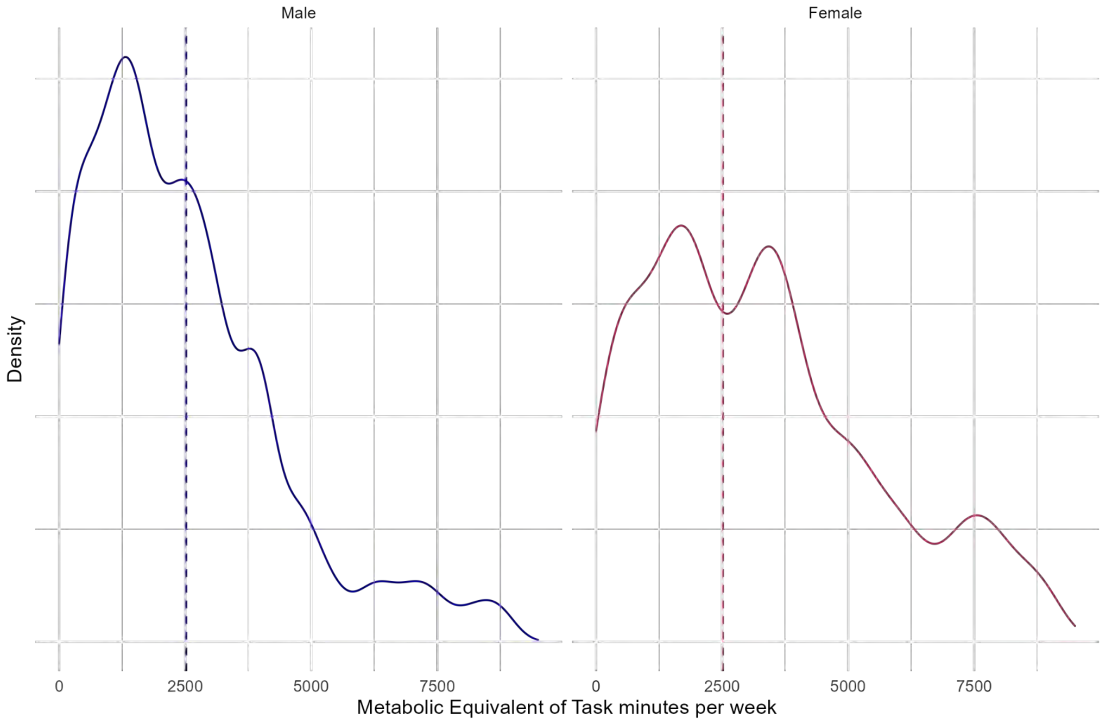
Data was obtained from the Health Survey for England (2015) and filtered to a dataset of 11- and 12-year-old respondents only. The dataset contains detailed estimates of the number of minutes of a large number of physical activities undertaken by each child in the past week alongside data on the child's sex and linked data to the quintile of Index of Multiple Deprivation (IMD) score. Children who reported

more than 7,000 minutes (16 hours per day) of moderate to vigorous physical activity in the previous week were assumed to have entered data incorrectly and were excluded. A compendium of metabolic equivalent of task (MET) estimates from Butte et al. 2018 was matched to each activity in the HSE dataset manually and used to estimate the total weekly MET minutes for each child in the sample (see Appendix A3 for the MET values used).

Sampling weights were estimated by 1) assuming that there are even numbers of male and female children in the population, 2) using ONS data on dependent children in each Lower Layer Super Output Area (LSOA) to estimate the target proportion in each quintile of IMD. The latter is necessary because children are more likely to live in areas with higher levels of deprivation than the general population. In particular children have a greater than 20% chance of being in the bottom quintile of LSOA IMD scores. Combining these two sources of information results in a weight for each individual child, which was used when sampling to generate a representative sample of dependent children in the general population.

A density plot of the initial levels of physical activity at age 10 are shown in the figure below. The dotted lines show the level of MET-mins that would be achieved if an individual followed guidance and participated in moderate to vigorous activity averaging 6 METs (equivalent to walking at a brisk pace) for an average of 60 minutes each day. The proportion to the left of the line (not meeting guidelines) is 57% for male and 41% for females at age 10 in the baseline simulation. This is higher than many of the estimates reported in Chapter 1, and the higher rates of insufficient physical activity in males in the data is surprising. The physical activity report from the Health Survey for England 2015 estimates that only 24% of boys and 18% of girls met PA guidance (Scholes & Mindell, 2016), but this cannot be compared to the proportions estimated from the distributions below because guidance at the time was for over 60 minutes of PA on every day without exception and does not include a direct estimate of MET minutes. Very few individuals have activity levels above 5000 METs per week, equivalent to averaging 60 minutes of distance running per day, which is as expected.

Figure 3 Initial physical activity levels, measured in weekly Metabolic Equivalent of Task minutes, for females and males in the simulation. The dotted line represents the threshold required to meet guidelines.

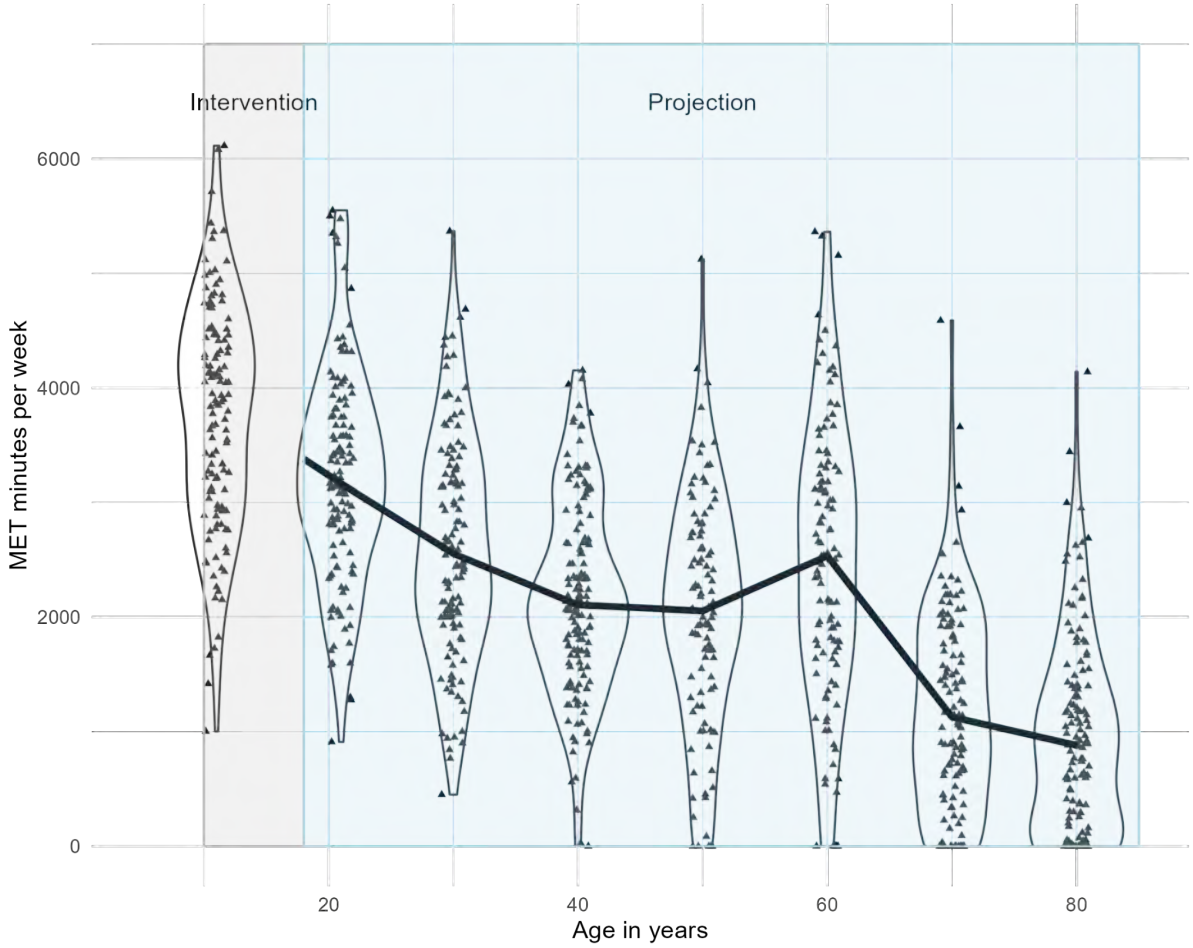


Trajectories of Physical Activity

Chapter 1 referenced the literature showing that physical activity tends to decline over the life-course (Bauman et al., 2012) and especially in adolescence (Dumith et al. 2011a). The model aims to capture this by splitting the simulation into two periods, from age 11 to 18 and from age 18 onwards. From age 11-18 trajectories of physical activity are informed by physical activity at baseline and a constant rate of decay of 6.1% (95% CI 3.9% - 8.3%) as per Dumith et al. 2011a. At age 18 in the *No Intervention* scenario, a cumulative density function is derived from the physical activity level of males and females separately. Each individual is assigned their percentile based on this distribution, and thereafter follow their percentile of physical activity until death such that an individual in the 45th percentile age 18 will still be in the 45th percentile age 70. This approach is based upon research indicating that PA tracks relatively well in adulthood, although trajectories are much less stable in childhood (Farooq et al. 2020; Lounassalo et al. 2019). The physical activity distribution for each age from 18 to 80 is estimated using self-report questions included in the Health Survey for England (HSE) 2014, 2015 and 2017. For the midpoint of each five-year age-band a set of physical activity percentiles were calculated by rank ordering the sample (split by sex) and cutting the sample at a sequence of cuts in increments from 0 to 1 in units of 0.01. Age specific estimates were derived using linear interpolation between the age-midpoint values for each percentile of the distribution. The intervention group are assigned a percentile from the baseline distribution based upon their activity level age 18 and, in the absence of effect waning, follow this percentile until death.

Figure 4 below shows a schematic of the method for a simulated dataset of females. The solid black line is the trajectory of the individual who had the median physical activity level at age 18 in the baseline scenario. Each year, her physical activity level changes such that she follows the linear interpolation between the median level of physical activity at the mid-point of each age-band. The benefit of this approach is that the *No Intervention* simulation maintains the currently observed distribution of physical activity at each age. The limitation of this method is that it assumes that there are no cohort effects, and therefore the median female aged 18 will have similar levels of activity when she is 75 to a 75-year-old today.

Figure 4 Visual representation of the method used to project physical activity levels for females from age 18 to age 80 (same method used for males separately). The solid black line shows the median female trajectory, measured in units of MET minutes.



NOTE: The schematic has been created to illustrate the method and is based upon a simulated dataset.

Intervention Effect

Size of Effect

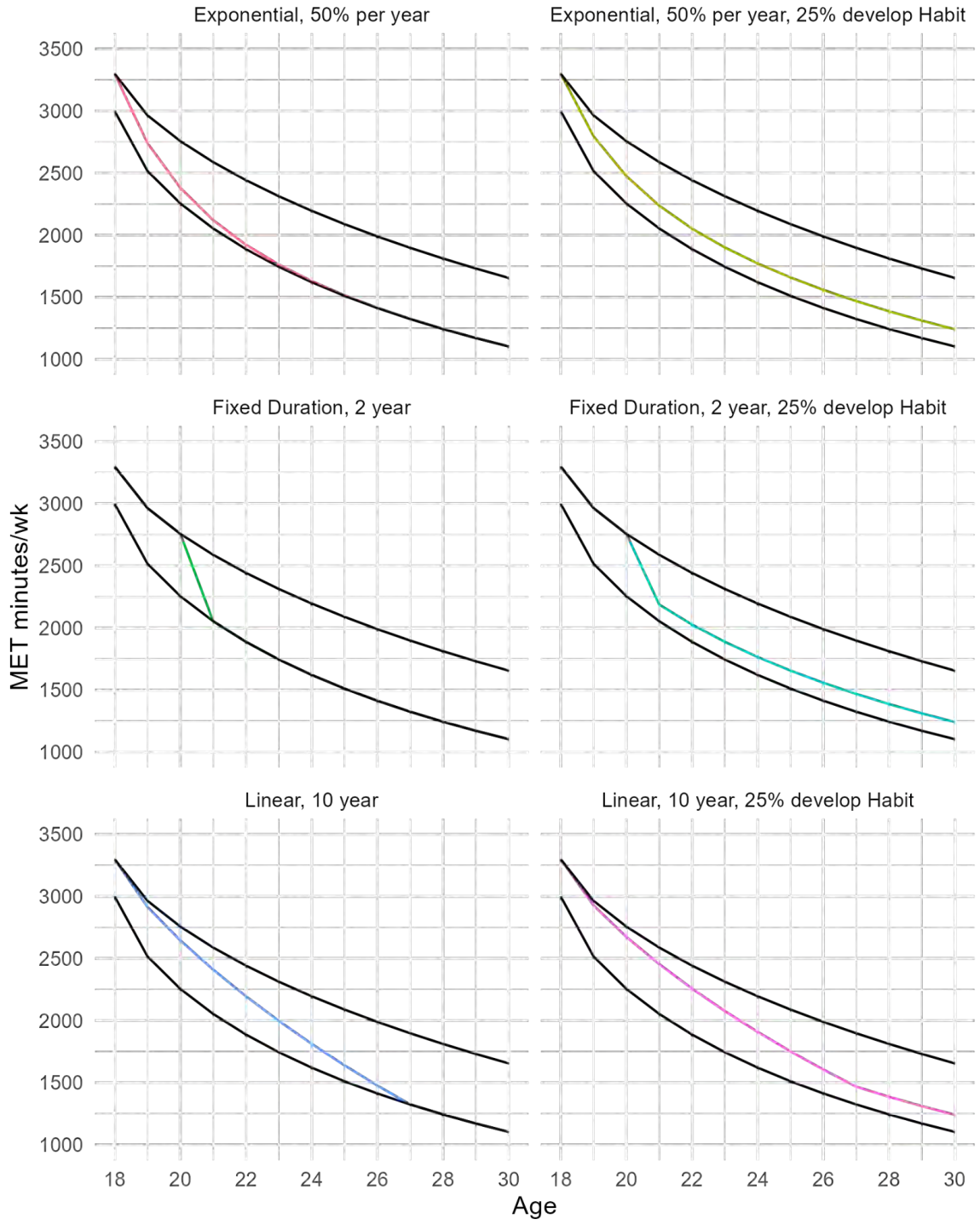
The case-study intervention is an after-school physical activity intervention described in Gc et al. (2019), which results in an additional 4.84 minutes (95% CI −0.94 to 10.61) of moderate to vigorous physical activity (MVPA) per day (Mears and Jago, 2016). This estimate of intervention effect is converted to a weekly MET-mins estimate by; assuming that the values reported are the average over a 7-day week, assuming that MVPA corresponds to a metabolic equivalent of task (MET) of 5.5, giving an intervention effect for the model of 186 (-36 to 408) weekly MET-mins. Each child is assigned an individual increment from this distribution, capturing the distribution around the mean effect. This intervention is assumed to occur for the period at which children are at secondary school (from age 11 to 18), at which point the intervention ceases and the effect decays. The intervention is assumed to have no direct impact on health-related quality of life or health-care costs. The indirect effect, via the five diseases and mortality, is modelled in the epidemiological model.

Decay of effect

There is no decay of intervention effect while the intervention is ongoing, which is assumed to be until age 18. At age 18, individuals are assigned to a percentile from the age 18 *No-Intervention* scenario physical activity distribution, so that if the intervention group is more active, they will be assigned a more active percentile than the no-intervention scenario group. From age 18 onwards, the intervention's effect on MET-min levels decays back towards the *No-intervention* scenario trajectory. This decay occurs at a constant rate of 50% in the central case, as assumed in Gc et al. (2019), such that the intervention group approaches the no-intervention scenario in around seven years. The effect of this common assumption is tested by comparing scenarios in which some proportion of the population (0%, 10%, and 100%) maintain the physical activity habit for the remainder of their life and scenarios in which the physical activity level decays linearly over 10 years. The 100% maintenance scenario is identical to the 0% waning scenario run by Gc et al. These figures are used to investigate the impact of structural assumptions in the model and do not reflect expert opinion of feasible parameter ranges. They are compared in a table and individually as scenarios on a tornado diagram.

Figure 5 below shows some of the options for modelling decay in effect, similar to previous approaches taken such as for a model of Type 2 Diabetes (Bates, 2021) In the absence of effect decay, an individual experiences the trajectory of physical activity associated with an individual at their level of physical activity at age 18, shown by the top black line in the figure. Three options for decay in effect size (Linear, Exponential and Fixed Duration) are shown as rows, and an option to allow a proportion of the population to develop a habit and experience no decay is shown on the right-hand column. Although the model is set up to enable a fixed duration effect, only sensitivity analyses using exponential and linear decay with a proportion of the population developing a habit are reported.

Figure 5 Decay in the mean effect of the intervention from the full intervention effect trajectory (top black line) to the baseline physical activity trajectory (bottom black line) for a hypothetical cohort of 10,000 individuals aged 18 under six different illustrative scenarios. The x axis (age) is cut at 30 to aid visualisation.



Lines as follows - Red: Exponential decay of 50% per year for entire population; Yellow: Exponential decay of 50% per year for 75% of the population; Green: No intervention effect after 2 years for entire population; Light blue No intervention effect after 2 years for 75% of population; Dark blue: Linear decay over 10 years for entire population; Purple: Linear decay over 10 years for 75% of population.

The use of a microsimulation model allows for much more flexibility in modelling approach - allowing intervention effects to be varied at the individual level. In particular, the ability to vary the intervention effect alongside the dynamics which would be expected to be observed as people age in the absence of the intervention is an underutilised modelling method. The approaches above are therefore different to approaches previously described for Markov models which make simplifying assumptions about probability of transition between physical activity 'states' over the lifecourse (Candio et al. 2021).

Epidemiological module

Risk of disease

The model focuses on five diseases that have a clear relationship between physical activity and risk, and for which there is a continuous dose-response function available: Breast Cancer (BC), Colon Cancer (CC), Type 2 Diabetes (T2D), Ischemic Heart Disease (IHD), and Stroke (IS) all of which are reported in Kyu et al., (2016). At the start of the model, all children are assumed to have none of the five diseases. During each cycle, individuals have a probability of developing each of the five diseases independently, and of dying, based on their age, sex, and MET-mins. For example the probability of a female with MET-mins of 600 being diagnosed with Breast Cancer between age 60 and 61 is calculated by multiplying the incidence of Breast Cancer in 60 year old women, obtained from the Global Burden of Disease 2019 database (Vos et al., 2020), with her relative risk of BC calculated from her physical activity level using the PA dose response function for BC (described below). As a result, more active people have a lower probability of developing health diseases than less active people. Individuals can develop multiple diseases simultaneously, and the risk of developing each disease is modelled independently - a weakness discussed further in the limitations section.

Incidence and prevalence of each disease were obtained for males and females in five-year age-bands from 0 to 90 (Vos et al., 2020). The central estimates and 95% confidence intervals around them were linearly interpolated from age 10 to 80 to derive single age-specific incidence (and 95% CI) for each sex and disease separately. When conducting probabilistic sensitivity analysis, a percentile for each disease was selected at random and used to extract a vector from age 10 to age 80 of incidence or prevalence. This approach is preferred to sampling incidence or prevalence each year independently since it is assumed that the uncertainty in the GBD estimates is likely methodological and therefore would be correlated between ages. We assume that there are no cohort effects, i.e. that independent of physical activity, the cohort of children will experience similar incidences of disease when they reach age 'a' to the incidences currently observed in the population aged 'a' .Trying to predict the long term trends in the incidence of each disease over a 70 year period was deemed to be beyond the scope of this study.

Dose response relationship with physical activity

The parameters used to estimate the relative risk function for each disease by weekly metabolic equivalent of task minutes are shown in Table 9 below. The log-linear relationships were originally derived from Kyu et al. (2016)'s estimates and are reported in Love-Koh et al. (2017)'s model described in Chapter 2. Uncertainty in the relationship was incorporated by sampling coefficients from a normal distribution informed by the 95% confidence intervals reported by Love-Koh and displayed in the table below.

Table 9 Coefficients for the log-linear model used to estimate the relative risk of five diseases from weekly metabolic equivalent time minutes (MET-mins) per week. The upper and lower 95% CI for the coefficients are shown separately.

Disease	Mean (SD)		95% CI Upper		95% CI Lower	
	Constant	ln(MET-mins)	Constant	ln(MET-mins)	Constant	ln(MET-mins)
IHD	1.016	-0.030	0.985	-0.034	1.043	-0.026
IS	1.073	-0.037	1.026	-0.043	1.140	-0.033
T2D	1.108	-0.040	1.108	-0.045	1.114	-0.035
BC	1.110	-0.024	1.120	-0.031	1.095	-0.017
CC	1.088	-0.031	1.076	-0.036	1.092	-0.024

Mortality rates

Disease specific mortality rates and other cause mortality was estimated from incidence, prevalence and mortality by age extracted from the GBD study database (Vos et al. 2020) using the DISMOD software (Barendregt et al. 2003), which assumes that mortality from all other causes is independent of disease specific mortality. Other-cause-mortality, the remaining mortality after these diseases are excluded, was derived by subtracting the disease specific mortality from overall mortality rates published in ONS life tables (Sanders, 2017). For individuals with multiple diseases, fatality rates are combined additively. The duration for which each disease impacted fatality rates differed by disease, with increased risk from oncological diseases lasting ten years, and metabolic and cardiovascular diseases being assumed to result in life-long higher rates of mortality.

Economic module

Intervention Cost

The intervention cost is included at £57 per individual, as estimated for a school-based teaching assistant-led extracurricular physical activity intervention (Jago et al. 2014), and inflated to 2020-21 prices using inflation indices from Jones and Burns (2021). The same source was used to inform

intervention costs in Gc et al., 2019. This cost includes training of teaching assistants to lead physical activity sessions, as well as some minor equipment and administrative costs.

Health State Utilities

The model estimates the HRQoL for each individual in annual cycles based on the disease status of the individual in that cycle. Individuals without any of the five diseases may still have other health conditions, but since the model includes five highly prevalent diseases it is likely that they would have a higher level of HRQoL than average. Therefore, quality of life in the absence of the five included diseases was estimated by taking the weighted average of general population HRQoL and the HRQoL for those with no reported health conditions (Ara & Brazier, 2011), where the weight was determined by an estimate of the proportion of overall health condition burden at each age that is attributable to the five diseases included in the model. The assumption used to inform the weight was that the percentage contribution of the five diseases to the difference between '*no reported health condition*' HRQoL and general population HRQoL increased linearly from 0% at age 11 to 50% at age 80. This assumption was based upon a cumulative proportion of total disease burden for the five diseases of around 30% observed in the Global Burden of Disease study (Vos et al., 2020), and the observation that the prevalence of the five diseases is disproportionately skewed towards older persons.

In the base case, a multiplicative approach is used, such that HRQoL is the product of the individual health state utility values. A further analysis is conducted using the minimum approach where HRQoL is equal to the minimum HRQoL of any of the individual health state utility values. These approaches of estimating quality of life in the presence of comorbidities are explained in more detail in Ada & Brazier (2010). Health state utilities for each disease were identified from Sullivan et al. (2011), as also previously used in the decision model described in Gc et al. (2019). Health state utilities are divided by the mean health related quality of life for the 60-70 age group in the model to give a set of utility multipliers. These multipliers are then combined multiplicatively, or the minimum used and multiplied by the estimated health state utility of the population with none of the five diseases. The uncertainty around these central estimates is incorporated into the model by using random draws from a beta distribution for each disease using the distributions shown in Table 10. Death is assumed to have a quality of life of zero.

Table 10 Health related quality of life mean value for illustration, and distribution sampled from in the model.

Disease	Health Related Quality of Life (Mean)	Distribution	Source
IHD	0.65	Beta(a=357,b=191)	Sullivan et al. 2011
IS	0.52	Beta(a=355,b=323)	Sullivan et al. 2011
T2D	0.66	Beta(a=5032,b=2548)	Sullivan et al. 2011
BC	0.76	Beta(a=791,b=256)	Sullivan et al. 2011
CC	0.67	Beta(a=150,b=73)	Sullivan et al. 2011

Health State Costs

The model assigns individuals an annual healthcare cost based on their disease status (see Table 11) and accumulates costs for individuals with multiple diseases. A single lifetime cost of treatment for Breast Cancer and Colon Cancer is assigned in the first year (following Gc et al. 2019), while cardiovascular diseases have a higher first-year cost and lower costs in subsequent years. We use a simplifying assumption of constant annual Type 2 Diabetes costs over the remainder of an individual's lifetime. The costs were based on data from a previous study (Gc et al., 2019) and were inflated to 2020-21 prices using inflation indices from Jones and Burns (2021).

Table 11 Health state costs in the first year and subsequent years with mean value for illustration, the distribution used in the model and the source of the data.

Disease	**Mean**	**Distribution**	**Source**
First year			
IHD	£6,187	Gamma (α=100, β=62)	Ward et al. 2007
IS	£11,110	Gamma (α=100, β=111)	Ward et al. 2007
T2D	£1,441	Gamma (α=100, β=14)	Clarke et al. 2003
BC*	£13,566	Gamma (α=100, β=136)	Madan et al. 2010
CC*	£18,830	Gamma (α=100, β=188)	Tappenden et al. 2007
Subsequent years			
IHD	£221	Gamma (α=100, β=2)	Ward et al. 2007
IS	£2,992	Gamma (α=100, β=30)	Ward et al. 2007
T2D	£1,257	Gamma (α=100, β=13)	Clarke et al. 2003

Discounting

Both costs and quality of life estimates were discounted at a constant rate of 3.5% per year, but this rate was varied in sensitivity analysis to 1.5%. The total costs and quality-adjusted life years (QALYs) for the intervention and no intervention scenarios were calculated for the entire cohort each year, with half cycle correction applied, and averaged across the iterations of the probabilistic model. All costs and QALYs are reported per-capita.

Incorporating Probabilistic Sensitivity Analysis and reporting results

To incorporate parametric uncertainty in the model, probabilistic sensitivity analysis (PSA) was run on the baseline model and all one-way sensitivity analyses (OWSA) as recommended by NICE guidance (NICE, 2014). 5,000 iterations were run for the central scenario and 2,500 for each OWSA. The results are shown to be stable at around 2,500 PSA iterations in Appendix A5. The distributions around central values for costs and health state utilities are provided above, to be consistent with the study by Gc et al. The duration of BC and CC was sampled from a triangular distribution (8 - 12 years) with a mean of 10 years. Health state specific mortality was sampled from a uniform distribution with a range 5% either side of the mean estimate from DISMOD. A dose response function was created each iteration using uncertainty around the estimates of coefficients provided by Kyu et al. 2016. Baseline disease incidences were sampled using the 95% confidence intervals from Vos et al. 2020 to inform the standard deviation around the mean value for each age. Physical activity percentiles at each age for the population were fixed.

The results are reported by way of an incremental cost-effectiveness results table and visualised graphically using a cost-effectiveness plane (CE-Plane), a cost-effectiveness acceptability curve (CEAC) created using the open source '*darkpeak*' package in R (Smith & Schneider, 2021c), and an economically justifiable cost analysis plot.

Sensitivity analysis

Sensitivity analyses were undertaken on the probabilistic model to estimate the impact of several methodological assumptions commonly made when building health economic models of physical activity.

- To assess the effect of habit formation, the proportion of the population that retain their higher physical activity tracking for the remainder of the life-course was varied from 10% to 0% and 100%.
- To compare the use of an exponential decay rate of 50% annually with a linear decay rate of 10 years for the intervention effect. We vary the assumption of multiplicative utilities with that of minimum utilities when combining disease utility decrements for multiple diseases.

- To understand the sensitivity of the model to the effectiveness of the intervention, the effect size was varied by 50% either side of the mean effect size (multiplying the entire individual level distribution by 0.5 and 1.5)
- To determine the effect of discounting, the discount rate for costs & quality of life was varied from 3.5% to 1.5%.

The table below shows the parameters used for each scenario. In each scenario only one parameter is edited from the central default model run at a time (One Way Sensitivity Analysis).

Table 12 Scenarios run as One Way Sensitivity Analysis in the model. Each structural or parametric assumption varied in any of the scenarios is included as a column, with each scenario given a name (first column).

Name	Proportion developing habit	Decay method	Utility method	Discount Rate	Number of iterations	Population	Int Effect (METmin / week)
Central (default) model run	10%	Exp	Mult	3.5%	5,000	10,000	186
0% develop habit	0%	Exp	Mult	3.5%	2,500	5,000	186
100% develop habit	100%	Exp	Mult	3.5%	2,500	5,000	186
Linear decay	10%	Linear	Mult	3.5%	2,500	5,000	186
Minimum HRQoL method	10%	Exp	Min	3.5%	2,500	5,000	186
1.5% Discount Rate	10%	Exp	Mult	1.5%	2,500	5,000	186
Lower Intervention Effect	10%	Exp	Mult	3.5%	2,500	5,000	93
Higher Intervention Effect	10%	Exp	Mult	3.5%	2,500	5,000	279

Mult = Multiplicative; Min = Minimum; Exp = Exponential

The results of the OWSA are visualised on a CE-Plane and a tornado diagram, and sensitivity analysis on intervention effect was displayed graphically on the economically justifiable cost-analysis plot.

Assessing external validity

The model structure and mechanics were critiqued by experts throughout the development process. Furthermore, while it is difficult to quantitatively assess the external validity of the mechanics of a model

constructed ex-ante, it is feasible to compare some outcomes of the 'No Intervention' model run to where data exists. The incidence and prevalence of each disease in the model population was compared with the Global Burden of Disease 2019 Incidence and Prevalence estimates (Vos et al., 2020), overall survival with life tables reported by the ONS (Sanders, 2017), and with estimates of Health-Related Quality of Life from Ara & Brazier (2011).

Model code validation and documentation

The model was programmed in the 'R' language (R core team, 2022). Where possible, the coding framework follows Alarid-Escudero et al. (2019) and is documented using *Roxygen* (Vidoni, 2022). The model code has been re-run independently by a third party to verify that the code achieves the same results, as recommended by the ISPOR-SMDM Modelling Good Research Practices Task Force (Eddy et al. 2012). The full model source code is available on GitHub at *https://github.com/RobertASmith/PACEM*. Data is available upon request.

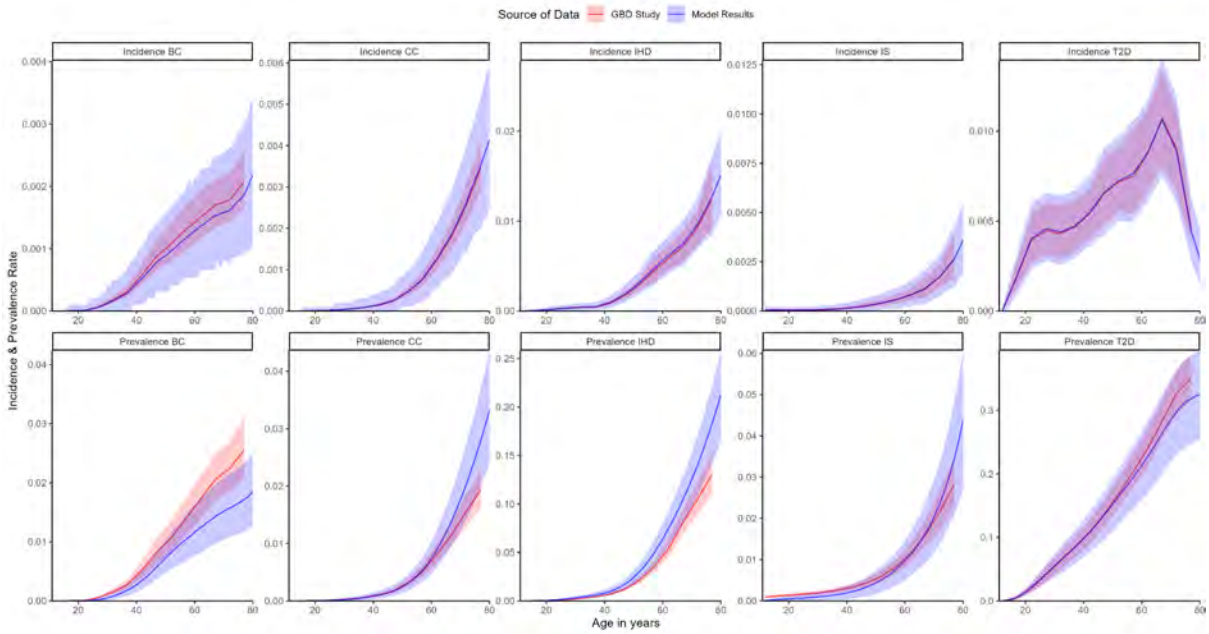
Results

There are three main sections to the results. The first section, *Model external validity*, compares the *No Intervention* model outcomes to data where they do exist. The second, *Model results*, reports the results of the model run using the central estimates for parameters and the default structural assumptions. The final section, *Sensitivity analyses,* runs extensive sensitivity analysis on the parametric and structural assumptions of the model.

Model external validity

The external validity of the base model was assessed visually against external datasets for the incidence and prevalence of the diseases, mean health state utility, and overall survival, by age. Figure 6 shows the mean and 95% CI for 5,000 iterations of the no-intervention scenario model run against the central and 95% CI of the Global Burden of Disease 2019 Incidence and Prevalence estimates (Vos et al., 2020). Incidence, as would be expected, exhibits a high degree of concordance, while prevalence, complicated by the incorporation of estimates of duration of disease and fatality rates, has lower levels of agreement.

Figure 6 Incidence and prevalence of diseases from age 11 to age 80 from the no-intervention model results (blue) compared to the Global Burden of Disease study 2019 for the general population in the UK (red). The mean is shown as a solid line, with 95% CI shown



Two additional visuals are shown below. Figure 7 shows survival in the no-intervention model run compared to the ONS life tables (Sanders, 2017). The model overpredicts overall survival in the lifetables, likely because of a combination of factors: projections of disease prevalence in the baseline scenario differ slightly from that observed in the GBD study. Updates to the model methodology after the calculation of other cause mortality have resulted in a mismatch. This should be addressed in future work.

Figure 7 Survival from age 10 in the model results, with 95% credible intervals, compared to ONS Life Tables.

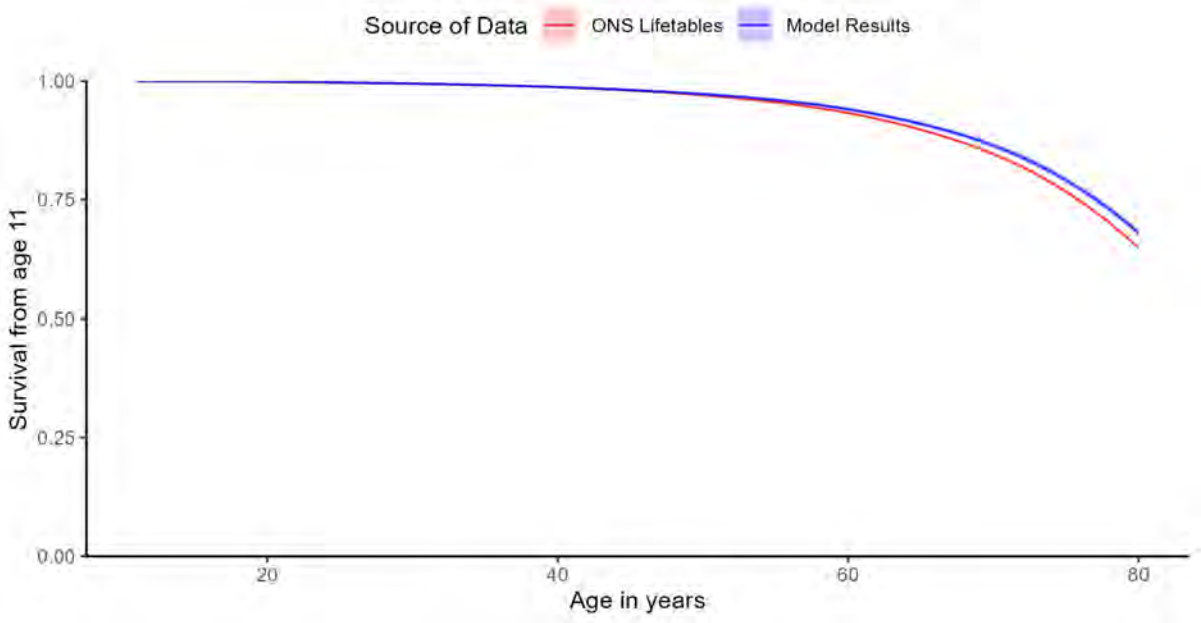
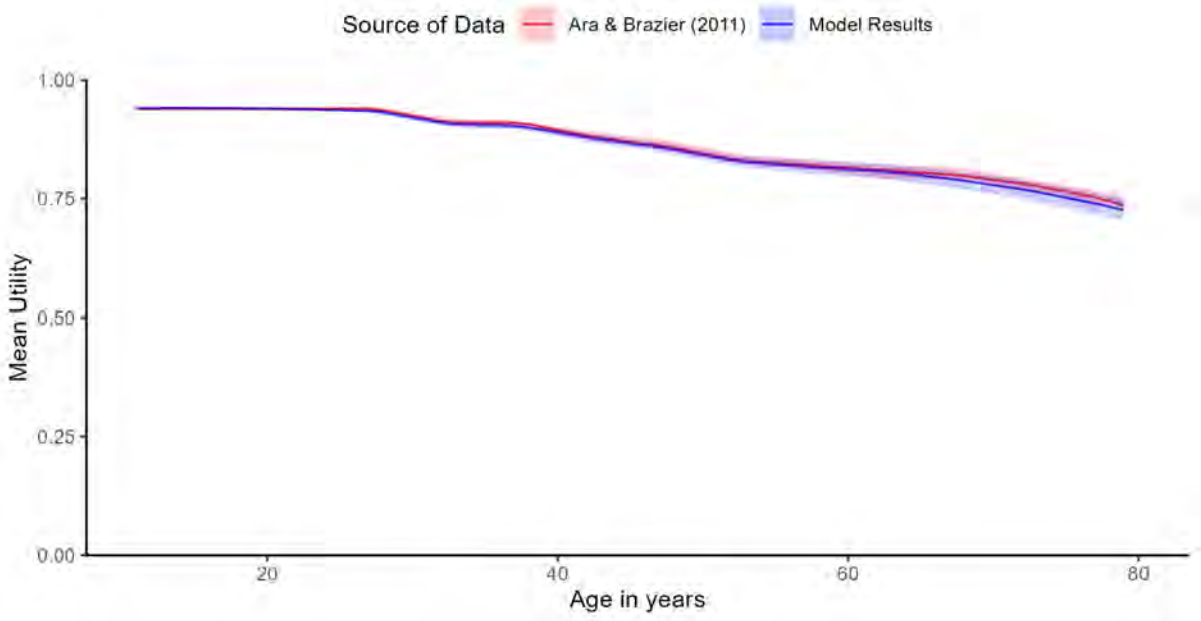


Figure 8 shows the health-related quality of life for the general population from Ara and Brazier (2011), compared to health-related quality of life for those alive in the model. For Ara and Brazier, we use linear interpolation between the mid-points of the reported age-groups for the mean and 95% confidence interval to generate a central estimate and 95% CI ribbon for each individual age to make the two outputs comparable. This compares favourably with considerable overlap between credible intervals.

Figure 8 Health Related Quality of Life for the general population (Based on Ara & Brazier 2011) compared to model results, from age 11 to 80.



Model results

The results of the analysis, summarised in Table 13 below, estimates that the intervention would result in a discounted incremental cost of £45 per person and increase discounted QALYs by 0.001 per person over the remainder of the life-course. The Incremental Cost-Effectiveness Ratio, the cost per Quality Adjusted Life year gained, is £37,438, which is above the willingness to pay threshold of £20,000. Therefore, the incremental net benefit is negative, at £21 per person.

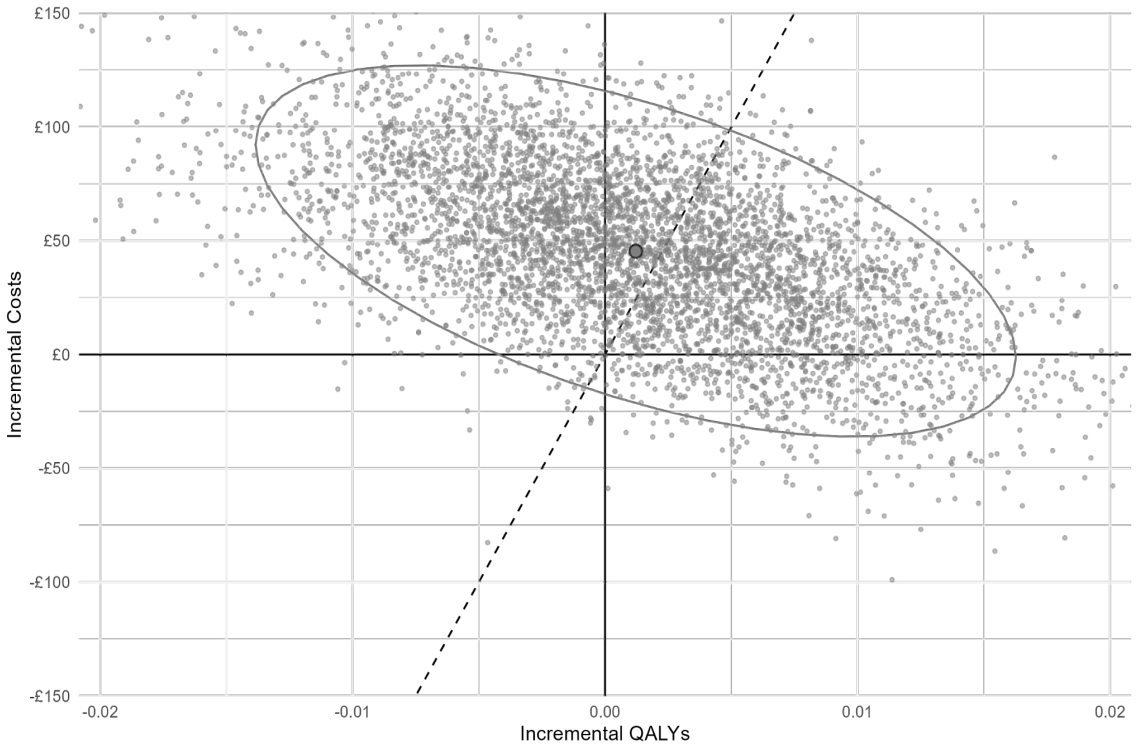
Table 13 Incremental Cost-effectiveness Table for the Intervention versus the No Intervention strategy.

Scenario	Total cost (£)	Total QALYs	Incremental cost (£)	Incremental QALY	ICER (£)	INB at £20,000/ QALY
No Intervention	3,853.18 (2,689.60; 5,171.91)	23.766 (23.600; 23.936)	NA	NA	NA	NA
Intervention	3,898.62 (2,741.18; 5,210.51)	23.767 (23.600; 23.937)	45.44 (-29.30; 119.89)	0.001 (-0.012; 0.015)	37,438	-21 (-342; 305)

Cost-effectiveness Plane

There is considerable uncertainty in the model results, as shown by the cost-effectiveness-plane in Figure 9 below which depicts all iterations of the PSA. The mean estimate falls in the north-west quadrant, meaning the intervention has positive incremental costs and QALYs. The point for the mean is to the north-west of the threshold shown by the dotted line, indicating that the central estimate of the incremental cost-effectiveness ratio is above £20,000.

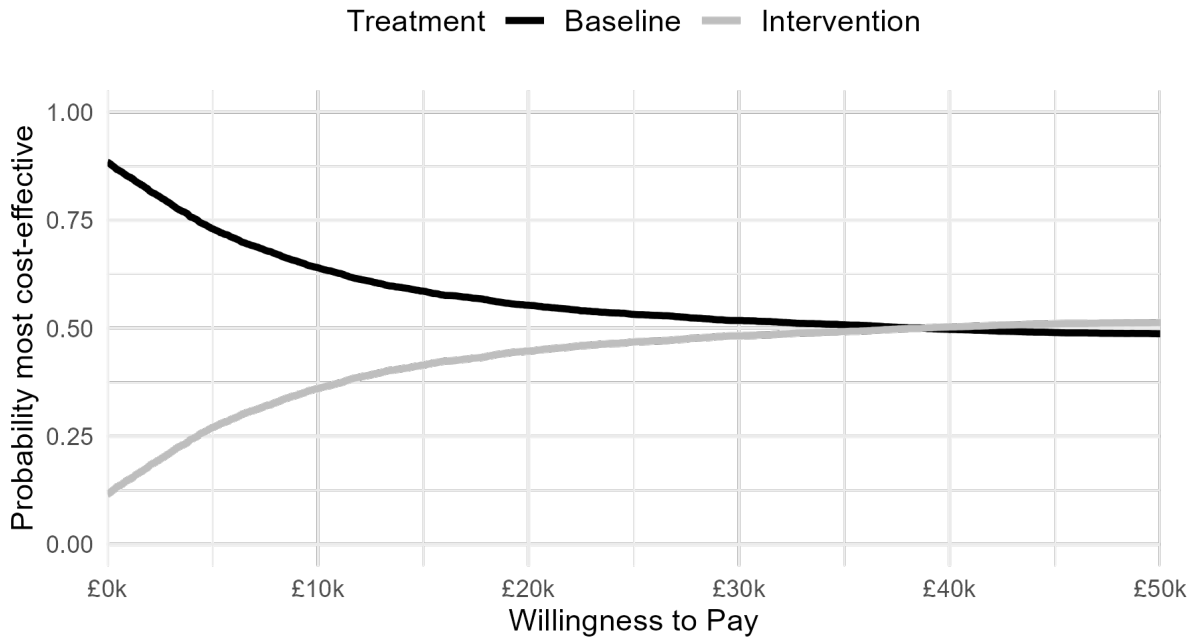
Figure 9 Cost-effectiveness plane for the base case at price £57 showing incremental costs and Quality Adjusted Life Years (QALY) for the Intervention relative to No-intervention scenarios. The dotted line represents a £20,000 threshold for a willingness to pay



Cost-effectiveness Acceptability Curve

A cost-effectiveness acceptability curve is shown in Figure 10 below. The intervention has a lower than 50% probability of being cost-effective at Willingness to Pay per QALY of £20,000, with the point at which the probability is greater than 50% corresponding to the ICER of £37k. An indication that the median resulting from the PSA results in the same finding as the mean. Furthermore, the less than 100% probability of cost-effectiveness even at a price of zero shows the impact of high levels of overall uncertainty in the model, and the small and uncertain marginal effect of the intervention.

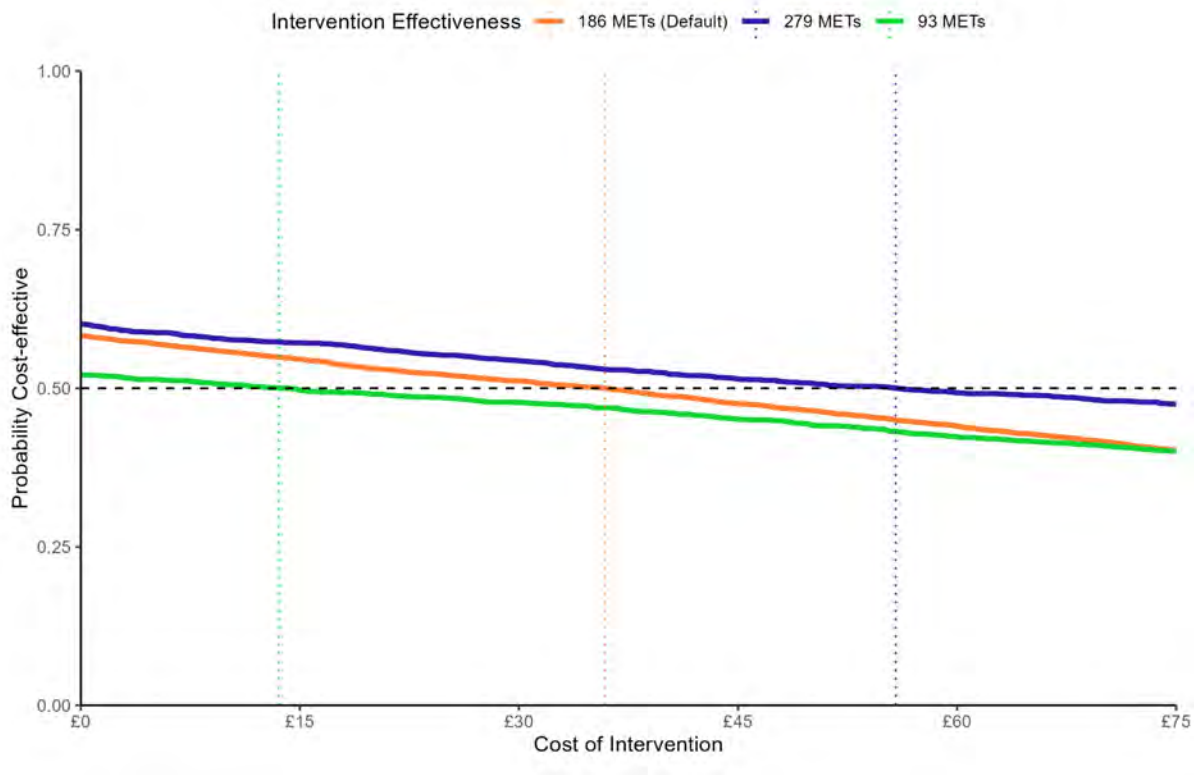
Figure 10 Cost-effectiveness Acceptability Curve showing the probability each strategy is Cost-Effective over a range of Willingness to Pay per QALY from £0 to £50,000.



Economically Justifiable Cost Analysis

Figure 11 shows the probability that the intervention is cost-effective for per capita costs between £0 and £75. Three scenarios are shown: the central scenario with intervention effect of 186 METs/week, a scenario with intervention effects 50% below the central scenario (93 METs), and a scenario with effect at 50% above the central scenario (279 METs). The cost at which there is a greater than 50% probability of cost-effectiveness differs by effectiveness scenario: at approximately £36 in the central scenario, £14 in the lower effectiveness scenario and £56 for the upper effectiveness scenario. Note that even at zero cost, because the model is probabilistic and the intervention effect is small and uncertain, the probability that the intervention is cost-effective is lower than 100%.

Figure 11 Probability that the intervention is cost-effective over a range of intervention costs from £0 to £75 at a willingness to pay per QALY of £20,000. Lines show, for three intervention effect sizes, the proportion of the PSA iterations for which the intervention is cost-effective. The dashed vertical line represents the price at which the intervention has a 50% probability (black horizontal dotted line) of being cost-effective. Note that the sensitivity analyses model runs (279 and 93 METs) have a smaller population size and so greater variance than the default model run.



Sensitivity analyses

The results of the OWSA, the methods for which are described in '*Methods-Sensitivity analyses* are visualised in the figures and tables below. Figure 12 shows the incremental cost and QALYs for each of these scenarios in comparison to a "No-intervention" group. Each iteration of the probabilistic sensitivity analysis for each scenario is represented by a small point coloured to represent the scenario. The mean result is shown by a large black triangle. In all cases there is considerable uncertainty about outcomes, reflected by a wide range of points. The assumption around long-term habit formation and discount rate is most influential to the cost-effectiveness estimates compared to the other sensitivity analyses. At an intervention cost of £57, the extent of habit formation had a significant impact on whether the intervention was considered cost-effective at a Willingness to Pay of £20,000 (as indicated by the dotted line), with the 100% habit formation resulting in the intervention becoming dominant (higher QALYs and lower cost). The choice of discount rate and the intervention effect size were also important. Smaller effects were seen for the choice of intervention effect decay function and HRQoL aggregation method.

Figure 12 Cost-Effectiveness Planes showing per capita incremental costs and Quality Adjusted Life Years (QALY) for each scenario relative to No Intervention on separate plots for each scenario. The dotted line represents a £20,000 threshold for a willingness to pay per QALY.

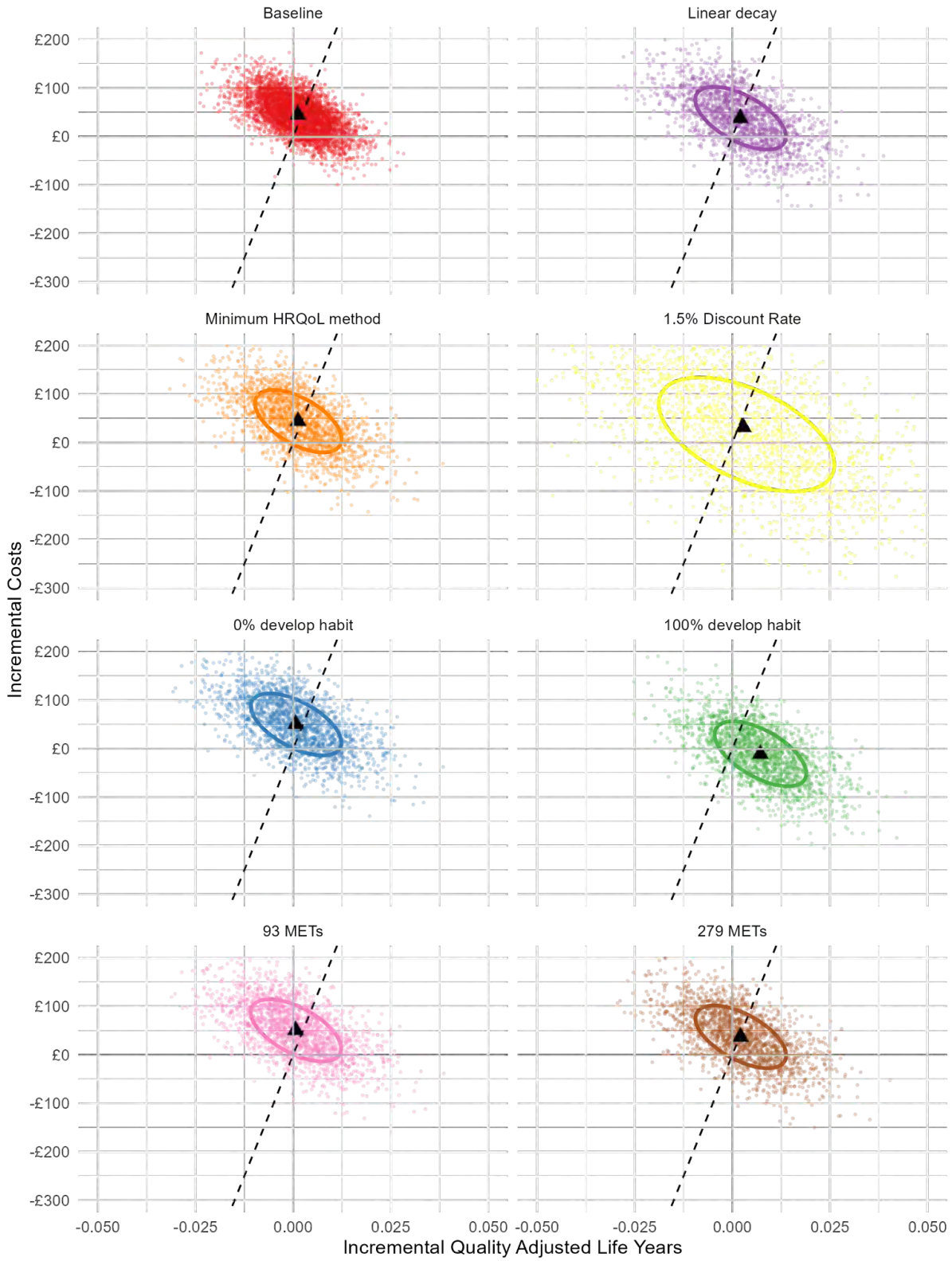


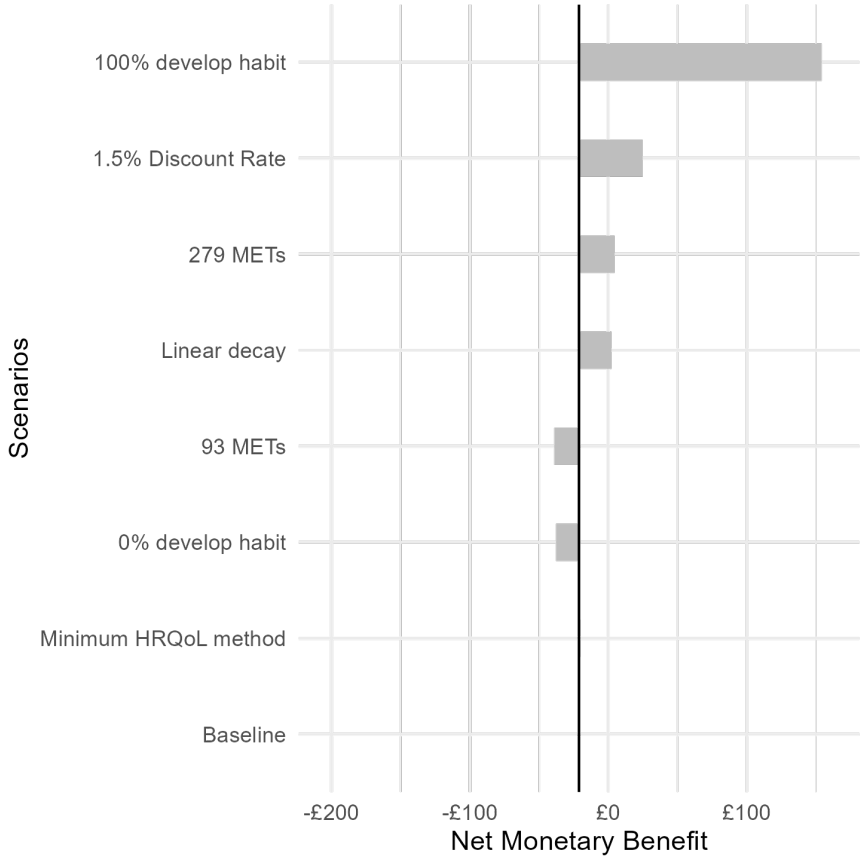
Table 14 Pairwise Cost-effectiveness Table for *Intervention* vs *No Intervention* for 5 sensitivity analyses: Habit formation, Intervention effect decay Function, HRQoL aggregation method, annual discount rate on costs & QALYs and intervention effect size.

Scenario	Total cost (£)	Total QALYs	Incremental cost (£)	Incremental QALY	ICER	NMB (£20k)
No Intervention	3,853 (2,690; 5,172)	23.766 (23.600; 23.936)	-	-	-	-
Intervention (Default)	3,899 (2,741; 5,211)	23.767 (23.600; 23.937)	45 (-29; 120)	0.001 (-0.012; 0.015)	37,438	-21 (-342; 305)
Habit Formation						
0% Habit (Pessimistic)	3,923 (2,712; 5,177)	23.765 (23.598; 23.933)	51 (-57; 161)	0.001 (-0.019; 0.021)	76,627	-38 (-511; 435)
100% Habit (Optimistic)	3,862 (2,666; 5,092)	23.772 (23.606; 23.939)	-10 (-125; 109)	0.007 (-0.013; 0.028)	Dominant	154 (-316; 644)
Intervention effect decay function						
10-year linear decay	3,911 (2,702; 5,154)	23.766 (23.600; 23.935)	38 (-70; 149)	0.002 (-0.018; 0.023)	18,854	2 (-464; 481)
HRQoL aggregation method						
Minimum	3,917 (2,706; 5,166)	23.788 (23.627; 23.950)	45 (-64; 157)	0.001 (-0.018; 0.021)	36,237	-20 (-474; 428)
Annual discount rate on costs & QALYs*						
1.5% *No intervention	9,256 (6,462; 12,117)	38.049 (37.687; 38.417)	-	-	-	-
1.5% *Intervention	9,288 (6,497; 12,149)	38.052 (37.688; 38.413)	32 (-202; 265)	0.003 (-0.040; 0.046)	11,252	25 (-951; 1,055)
Intervention effect size						
93 METs	3,924 (2,710; 5,181)	23.765 (23.597; 23.933)	51 (-57; 161)	0.001 (-0.019; 0.021)	83,014	-39 (-515; 432)
279 METs	3,910 (2,702; 5,153)	23.766 (23.599; 23.934)	38 (-71; 143)	0.002 (-0.018; 0.023)	17,843	5 (-466; 477)

**The No intervention costs and QALYs for sensitivity analyses differ for No-intervention when the discount rate is changed, therefore both intervention and no intervention results are shown.*

The results above are depicted on an adapted tornado diagram, showing each scenario on a line rather than minimum and maximum values for each, to visualise how Net Monetary Benefit varies among the scenarios.

Figure 13 Tornado diagram showing the Net Monetary Benefit for each scenario compared to the default scenario (vertical black line). The scenarios are ranked in order of variance from the default scenario Net Monetary Benefit.



Discussion

In a recent review of the methods used in models of physical activity, Candio et al. (2020) recommended that "*A more structured and formal approach to assessing the implications of key structural assumptions on the economic decision should be a minimum requirement. In absence of data, a range of plausible scenarios should be explored, and results be presented as such, rather than just a base-case*" (p.1160). While this is the case for all health economic models, the authors identified the lack of sensitivity analysis as being particularly acute in models of physical activity. This study has attempted to analyse the sensitivity of results to some of these key structural assumptions, in the context of an intervention which increases physical activity in children and adolescents resulting in habit formation for some. It builds on the work of Gc et al. 2019 in estimating long term costs and QALYs from interventions in

childhood and adolescence. The results of this study, which in the central scenario find a ICER of £37,438, which is higher than that of Gc et al. at £11,486. However, the two models have several differences, for example this model uses a new model structure (microsimulation model) which is better suited to capture individual level risks and physical activity trajectories, includes a proportion of the population for whom the effect did not decay, and includes an individual level non-linear dose response relationship between physical activity level and disease risk.

One large difference between this model and that of Gc et al. (2019) is that this model does not include any direct effect of PA on HRQoL. The assumption made by Gc et al, that the observed relationship between PA and HRQoL is causal, is likely to have a large impact on model results since 1) the effects are immediate, occurring in adolescence, 2) the effect size is relatively large (a 4pp increase in HRQoL) and 3) it applies to the entire population with increased PA, not just those who go on to develop a disease. Unfortunately, no sensitivity analysis was undertaken on this direct effect, but it is likely that in the absence of this effect the results would have been considerably less favourable than the results of the analysis described in this paper.

Sensitivity analysis - implications for model design

The results of the analysis described above indicate that the model is particularly sensitive to the percentage of the population that maintain increased PA over the rest of their lifetime. When assuming 100% of individuals form such habits (i.e. no decay in effectiveness), the cost-effectiveness of the intervention improved significantly compared to assuming 10% (the central case), making the intervention cost-saving and therefore dominant.

A recent review found that 15 out of 25 PA models in the published literature assumed no decay in the intervention effect for 100% of the population (Candio, 2020). Of the 10 that did assume some decay, three performed sensitivity analysis on the rate of decay (Roux et al. 2008; 2015; Gc et al. 2019; Cobiac et al. 2009). Roux 2008 does not provide much information, simply stating that "*varying the dissipation of the effect sizes of the interventions had a marginal impact*" (p584). Cobiac et al. 2009 and Gc et al. 2019 both varied the rate of decay from their central estimates (50%) to 0% and 100% as sensitivity analysis for their analysis of the cost-effectiveness of multiple interventions. Between the two studies, decreasing the central estimate of decay to 0% resulted in one or more interventions becoming cost saving, and increasing it to 100% resulted in one or more interventions exceeding the cost-effectiveness threshold. Another model by Frew et al. (2012) includes a scenario in which physical activity is reduced after one year for half of those who had improved their physical activity levels in a trial of community based physical activity programs for adults, in this scenario the ICER increased from £400 to £1,840, implying a significant (x4) effect on incremental QALYs, albeit with limited impact on the decision due to the low initial ICER. The sensitivity of previously published models, and the model described in this chapter, to assumptions about habit formation and rate of decay suggests that other models which

assume no decay in intervention effect are likely overestimating the cost-effectiveness of PA interventions in adolescents.

The model was also sensitive to the discount rate selected. This is because most costs and HRQoL benefits occur later in life. Decision makers should consider that many of the benefits of interventions for children will not be realised for a long time. A discount rate of 3.5% means that benefits accruing at age 60 for a 10-year-old are only worth a sixth of immediate benefits, potentially leading to short-sighted policymaking that prioritises '*quick-fix*' interventions to temporarily increase physical activity while under-investing in interventions which change long term habit formation.

The results were sensitive to both the cost and marginal effect of the intervention. Doubling the effect of the intervention meant that the justifiable cost could be increased by around 50%, while halving the effect size reduced the justifiable cost by around half. It is notable that even at zero cost the intervention was not certain to be cost-effective, since there is considerable uncertainty about both the benefits of the intervention and the intervention had a small marginal impact on costs and QALYs in the context of large uncertainties in many of the model parameters.

The model results showed low sensitivity to the method of aggregating health-related quality of life and the choice of decay function (exponential or linear) of the intervention effect. The decay function may be of low significance for this model because the benefits of physical activity occur far in the future, unlike other physical activity models which may be more sensitive to this parameter.

Strengths

There are several strengths of the underlying model that make it a valuable contribution to the field. Firstly, the model is a microsimulation that captures many individual level dynamics that Markov models cannot, in particular capturing duration of time in each of five independent health disease states simultaneously without the need for a large number of tunnel health states as in many previous studies (Gulliford et al. 2014). In this regard, it is a 'cleaner' modelling method than many of the previous models created using a Markov structure more common for pharmacoeconomic modelling.

In the absence of a whole-of-life longitudinal study, the approach taken here estimates the long-term impact of physical activity in a new way, capturing both the long-term trajectory of physical activity and the potential for habit formation and intervention effect decay over the life course. This is a simplified version of the method used in the Population Health Model (POHEM) developed by '*Statistics Canada*'. This is a relatively complex problem that has been simplified crudely in many studies in the past (Candio et al., 2020). This study therefore aims to abide by the recommendations by Candio et al. (2020) that "Natural trends in PA levels should be used as baseline data for comparison", and that "Assumptions regarding time lags and decay of intervention effects over time must also be made explicit" .

The model incorporates a non-linear dose-response relationship between physical activity and multiple diseases over the entire life-course. This allows for a more externally valid representation of the relationship between physical activity and disease outcomes. The model also incorporates a consideration of both decaying intervention effect, risks of disease and death based on physical activity history, and estimates health state utilities given combinations of diseases and other conditions not included in the model but assumed constant in the population.

As a result of these methodological improvements, the base model has been seen to result in similar incidence, prevalence, and health state utilities to that observed in epidemiological studies (Vos et al., 2020). Many earlier health economic models did not undertake this comparison, making it hard to determine their external validity. While this does not ensure that the marginal effect is accurate, it does at least indicate that effect sizes are being calculated on the correct disease burden in the absence of the intervention.

The analysis also has several strengths. Firstly, each structural assumption is evaluated within a probabilistic model which gives some indication of the impact on both the mean and the variance in results. Multiple structural assumptions are compared, and their impact contextualised against each other. Finally, all model code and outputs are made openly available for others to run.

Overall, the model and analysis meet the minimum standards identified by Candio et al. (2020) as limitations in many other models; modelling of downstream disease risks, using natural trends in PA as the baseline for comparison, explicit assumptions around decay of intervention effect, extensive scenario analysis and individual level differences in risks and outcomes.

Limitations

The current study has several limitations that should be considered when interpreting the findings. Firstly, the study only includes a cohort of a single age, which is routine, but some interventions may be targeted towards broader age-cohorts (O'Mahony et al. 2015). Secondly, while the approach taken to model long-term trajectories of physical activity is simple to understand and generally maintains credible population physical activity distributions at each age, it fails to account for individual variability in physical activity levels throughout the lifespan and does not account for cohort effects.

The model only has a time horizon up until age 80. As a result, the survival benefit of physical activity may not be fully incorporated into the model if those who are more active have better long-term survival past age 80. It is not clear how this will influence the model results, since, as discussed in Mytton et al. (2017) there are substantial healthcare cost increases associated with increased longevity which may result in a longer time horizon reducing the cost-effectiveness of the intervention.

Some diseases which are likely to be influenced by physical activity, discussed in Chapter 1, were not included in the final model (Warburton et al. 2017). Including mental health conditions and injuries in

childhood and adolescence may impact the model results substantially, as might incorporation of Dementia and Alzheimer's in older age - although it is worth noting that it is not always clear in what direction since increased longevity due to physical activity may greatly increase the cost of these conditions (Mytton et al. 2017). However, the model is programmed in such a way as to easily incorporate new conditions, so the hope is that others may be interested in building upon the methods described here. Additionally, the model does not include feedback loops between diseases and physical activity, or incorporation of competing risks between diseases, a common limitation of models in the field (Candio et al., 2020).

Furthermore, estimating the HRQoL, risks of death, costs, and other outcomes for combinations of diseases is fraught with challenges. We use the DISMOD calculator to estimate the disease specific mortality rates, but this assumes that the entire difference in mortality rates by disease status can be attributed to the disease. We have generally assumed independence between diseases with multiplicative risks and have combined HRQoL for comorbidities multiplicatively as per the NICE reference case (NICE, 2014). Child specific HRQoL are also not included in the model, since the bulk of the quality-of-life burden occurs in later life. Direct impacts of physical activity on HRQoL and healthcare costs were excluded, assuming to only occur via the diseases included, which may have led to an underestimate of the benefits of physical activity. However, this decision was made to avoid overestimating the benefits (via double counting) as is likely the case in previous studies. Costings are aggregated across diseases with each disease treated independently, and only allowed to vary in the first year versus subsequent years, rather than including information about trajectories of long-term costs, which may be important for diseases such as Type 2 Diabetes where costs are expected to change over time.

Finally, the model is slow to run, which makes it computationally challenging to perform additional sensitivity analysis and limits its usability in a user-interface with stakeholders.

Future Research

To address the limitations of the current study, future research could incorporate more diseases into the model, in particular depression, anxiety and injuries. Incorporating some of the feedback loops between different conditions included in previous models of diabetes (Breeze et al. 2017), such as increased risk of IHD after T2D, would greatly improve the external validity of the model, as would the incorporation of methods to calibrate disease fatality rates to better fit to observed disease prevalence. Additionally, future research could attempt to incorporate statistical analysis of HRQoL that includes covariates for age, sex, physical activity and comorbidities, and to include child specific HRQoL and cost estimates. Finally, efforts could be made to optimise the model to make it faster to run in order to be able to more rapidly perform additional sensitivity analysis and to be able to run it in real-time for stakeholders, considerably improving the usability of the model.

Implications for policy

Many of the benefits of increased physical activity, such as improvements in health-related quality of life (HRQoL) and reductions in healthcare costs, are disproportionately accrued in later life. This means that the largest effects, in terms of healthcare costs and HRQoL, of any changes in social norms and behaviours in young people over the past decade will not be observed for many years. Policymakers should take a long-term view when making investments aimed at increasing physical activity in younger populations. This is important not only to improve outcomes observed in older populations now, but also to mitigate the potential negative impact of changes in behaviour currently being observed in young people now, but which won't be visible as changes in healthcare costs and quality of life for many years.

When interpreting the results from health economic models of physical activity, decision makers should be particularly sensitive to assumptions around habit formation and the rate at which intervention effects decay. Models that assume no decay in intervention effect for any proportion of the population are likely to significantly overestimate the effect of interventions, particularly when longer term studies show that behaviour is often not maintained. Since investments in interventions that have low decay rates and high probabilities of long-term habit formation are more likely to be cost-effective, identifying activities that can be maintained or adapted over the life course should be a priority. These activities may differ by sex, socioeconomic status, and ethnicity.

Conclusions

This study has identified that several structural assumptions commonly made in health economic models of interventions aimed at increasing physical activity in children and young people significantly affect the outcomes of cost-effectiveness analyses. Most importantly, assumptions around intervention effect decay and habit formation are crucial. Undertaking sensitivity analyses within modelling reports and allowing stakeholders to vary these structural assumptions would help improve the robustness of model results.

Part 4 - New methods to improve usability, efficiency, and transparency of health economic models.

This section of the thesis focuses on how new methods in data science and computing can enhance the usability, transparency, and efficiency of the development and updating of health economic models.

Chapter 7 describes web-based user interfaces for health economic models and outlines their value, and the necessary skills for building them. It includes a peer-reviewed tutorial paper that has been used by various health economists, including those working with the World Health Organization, UK Health Security Agency, and several pharmaceutical companies, to guide the development of user interfaces (Smith & Schneider, 2020a). It also includes a discussion of a case study that applied these methods to make a geospatial public health decision model more accessible to decision makers, illustrating the value of web-based user-interfaces in helping engage and empower decision-makers (Smith, Schneider, Bullas et al. 2020b; Schneider, Smith, Bullas et al. 2020).

Chapter 8 expands on the previous chapter by proposing a method for the automation of health economic model updates that rely on sensitive data, while allowing the data-owner (e.g. a company, NHS trust or government department) to share data. It includes a peer-reviewed tutorial paper that provides a step-by-step guide to the code of an open-source prototype application (Smith, Schneider & Mohammed, 2022a). The method can be applied to any situation in which a health economic evaluation needs to be updated regularly and is not specific to models of physical activity interventions. However, a discussion of potential use cases specifically for models of physical activity is included in Chapter 8. This method has been presented at the EARL conference in London, R-HTA in Oxford, R-HTA-LMIC (Remote) and short courses provided to health economics teams within pharmaceutical companies, market access consultancies and a workshop organised with NICE. Stakeholders have suggested it is likely to be most useful where data is updated regularly, and where decisions have to be made quickly, with obvious contenders being infectious diseases. There was also a recognition that the method complements the increasing use of real-world evidence in health economics. However, it has not yet been applied in a real-world project.

Chapter 7. On the creation of web application user interfaces for health economic and public health models.

"*Authors of modelling papers can do things to make their work substantially more useful for policy. The best is to provide an interactive interface, where if the policymaker does not agree with the starting assumptions of the model, they can change them*." (Whitty, 2015)

Health economic decision models play a crucial role in decision making by those responsible for allocating healthcare resources and determining public health policy (NICE, 2014). However, as discussed in previous chapters, there is often a trade-off between the usability of a model, how easy it is for decision makers to use the model to improve their understanding and inform their decision, and the external validity of a model, the extent to which the model attempts to capture the complexities of the real world. In a previous chapter, I introduced the Production Possibility Frontier (PPF) to describe this trade-off and identified several factors that may result in the PPF shifting outwards. One of the factors identified was "*improvements in technology which improve the user-friendliness of the model's interface".* This chapter is about one of the technological advances in data science which can help to mitigate this trade-off by making it easier to create user-friendly interfaces for health economic evaluation models (Chang et al. 2022).

An interactive interface for a health economic model is a tool that allows users to interact with the model, input data, and view results. The design of a user interface can vary depending on the specific needs of the model and the intended user group, but some common elements include:

- Input fields to allow users to enter data into the model, such as patient demographics, treatment costs, and outcomes data.
- Output displays to show the results of the model, such as cost-effectiveness ratios or budget impact estimates. These may be presented in the form of tables, graphs, or other visualisations.
- Navigation to allow users to move between different sections of the model.
- Help and documentation to help users understand and use the model, such as a user manual or online help pages.

The result is a user-friendly and intuitive interface which enables users to easily input data, adjust parameters, view the results, and 'play with the model'.

There are many benefits of web-based user interfaces for health economic models as compared to user-interfaces contained in spreadsheet software:

- Accessibility: Web-based interfaces can be accessed from any device with an internet connection, which makes it easier for users to access the model from anywhere (Chang et al., 2022).
- Collaboration: Web-based interfaces allow multiple users to access and collaborate with the model simultaneously, which can be useful for team-based projects, or for expert elicitation (Williams et al. 2021).
- Ease of use: Web-based interfaces often have a user-friendly design, which makes it easier for users to navigate and use the model, even if they have little technical expertise.
- Scalability: Web-based interfaces can handle large numbers of users and data, making them well-suited for large-scale projects. There is a significant administrative burden of sharing large spreadsheet files between different groups of users who must all have that spreadsheet software installed that is alleviated by developing a web-based interface.
- Security: Web-based interfaces can be configured to have secure login and data protection features, which can help to ensure the confidentiality of sensitive data and/or intellectual property.
- Version control: Web-based user interfaces can be easily updated and modified remotely, which can be useful for making improvements or fixing errors in the model. These updates can be made quickly and efficiently, without the need to distribute new software to users. This ensures that the model being used is the most up-to-date version.
- Data updates: the data used by web-based models are typically stored on remote servers which can be kept externally valid, up-to-date, and secure. This can be especially important for models that rely on data that changes frequently, such as data on treatment costs or effectiveness, or data that is commercially sensitive (Hart et al. 2020).
- Computing power: web-based user interfaces run model code on a remote server with more powerful hardware and processing capabilities than that of personal computers. Incerti et al. 2019 note that "*methodological and computational advances now allow for models that are increasingly sophisticated and realistic but cannot be reasonably computed in Excel*" (p.575).
- Tailored analysis: As discussed by Incerti et al. (2019) "*decision makers can tailor analyses to their local population by modifying the characteristics of the target population or using parameters based on data relevant to the local setting*". This can be much more efficient than building a model for each setting.
- Transparency: Health economic models are typically presented in reports that include results and sensitivity analysis. However, these reports are static and do not allow decision makers to interact with the model or update inputs (Cohen & Wong, 2017). User interfaces can enable decision makers to change inputs and re-run sensitivity analysis, empowering them to engage with the model and better understand the impact of different assumptions. This is particularly important in public health, where there is often more uncertainty and heterogeneity in methods, and subgroup analysis is more regularly applied.

However, to build a web-based user interface for a health economic model, the following skills may be necessary:

- Health economics: A strong understanding of health economics and the specific need of the model is essential to design an effective and user-friendly user interface.
- Programming: Building a web-based user interface for a health economic model typically requires that the model is built in a programming language such as R or Python. Even those in the industry with experience "*acknowledge that there are difficulties associated with learning a script-based language for modelling*" (Baio & Heath, 2017). In addition, expertise is needed with building user-interfaces with software packages such as shiny (for R) or Dash (for Python) as well as programming languages such as HTML, CSS, and JavaScript if further customisation is required.
- Data management: The user interface may need to handle large amounts of data, so skills in database management and data manipulation may be necessary.
- User experience design: Building a user-friendly and intuitive user interface requires a strong understanding of user experience (UX) design principles.
- Project management: Building a web-based user interface may involve coordinating with a team of developers, designers, and other stakeholders, so project management skills may be necessary.

Given that each of these skill sets require years of training it is unlikely that all will be undertaken by a single individual, therefore project management skills may also be useful in coordinating the development process with multiple team members. Given that the health economic model is the core of the work, and typically requires the longest training of any of the skillsets (typically an MSc or PhD), and is arguably in shortest supply (Kaambwa & Frew, 2013), the most efficient way to upskill the industry is likely to give health economists the broad skillset to manage the process of building web-based user-interfaces for their models - drawing on the expertise of others where required.

One of the barriers to the proliferation of web-based user interfaces for health economic models was a lack of instructional material to provide experienced health economists with the necessary skills to deploy such applications. To address this issue, I worked with colleague Paul Schneider to create a tutorial paper on using the shiny software package in R (Smith & Schneider, 2020a) to build and deploy web-based user interfaces for health economic evaluation models. The paper, which has been published in an open access journal with publicly available reviewer reports, is the first to outline the method for constructing a user interface for cost-effectiveness analysis and to provide all the necessary code open source. The tutorial paper describes a method in which a health economic decision model, including probabilistic sensitivity analysis, is wrapped into a single function in R and controlled by numeric, slider, and drop-down inputs in the user interface. When the "run-model" button is clicked, the model function is executed on the server side, with user inputs as parameter values. The outputs of the model, such as plots and tables, are then displayed to the user.

While the paper is the first to outline the method of creating a web-based application for a health economic model, previous papers have described the functionality of applications to run Network Meta-Analysis (Owen et al. 2019; Baio et al. 2017) and Value of Information Analysis (Strong et al. 2015). A recently published paper also outlines the core functionality and provides a case for the use of R-Shiny in health economics more widely (Hart et al. 2020).

A PDF of the published tutorial paper is provided below, with a wider set of tutorial materials available on an online repository linked in the paper, with the aim of making these materials easy to access, which is essential, especially for countries where fewer opportunities for formal taught courses exist (Frew et al. 2018) . Since the tutorial paper was published, the method has been used to build and deploy applications for various organisations, including for the World Health Organisation Department of Reproductive & Maternal Health (WHO, 2022), parkrunUK (Smith et al. 2021b; Schneider et al. 2020; Smith et al. 2020b), the UK Health Security Agency and several companies in the pharmaceutical industry.

The work with parkrun illustrates the value to decision-makers of access to a well-designed UI for a model. A body of work undertaken with colleagues from ScHARR and Sheffield Hallam in 2020 led to the development of a model which was used to identify the optimal locations for new parkrun events in England (Smith et al. 2021b; Schneider et al. 2020; Smith et al. 2020b). However, the results were hard to articulate in a written report for laypersons due to the quantity of data input into and output from the model, and the complexity of the modelling methods. Through the deployment of a web-based user interface, currently hosted by the University of Sheffield (http://iol-map.shef.ac.uk/), the team at parkrun were able to better engage with the model. The web-app was used by the parkrunUK team to identify potential sites, which were then visited by volunteers on the ground to assess practical suitability. Sites which were deemed suitable were then designated as target sites and local links used to encourage the establishment of a new event. The value of this tool was outlined by parkrun in their blog, and I joined the parkrun podcast to discuss the methods (Parkrun Blog, 2020). Since none of the parkrun team had experience programming in R it was unlikely that they would have been able to benefit from the work in the absence of the web-based user-interface (UI). The final step of developing the UI, which took significantly less time than the conceptualisation of the problem or the development of the statistical model, meant that the work was able to directly engage decision makers and therefore inform policy. The argument, made repeatedly throughout this thesis, is that alongside methodological improvements to the underlying model, development of UIs can help build confidence in models by allowing users to adapt parameters based on new or different information, or to better understand the sensitivity of findings to those parameters. This approach simplifies complex statistical analysis, making it easily comprehensible for decision-makers who may not possess the necessary time or expertise to conduct such analysis independently. The next section, in the form of a published paper, provides an overview of how to build user-interfaces using open-source software in the R software environment.

A tutorial on the creation of user interfaces for health economic models

This section of the chapter contains a publication which outlines a method by which user-interfaces can be developed to improve the transparency and usability of health economic evaluation models.

This article was published open access following the requirement of the Wellcome Trust who financially supported this work. The conditions of the open access publishing allow use of the final published PDF, original submission or accepted manuscript in this thesis (including in any electronic institutional repository or database). The original content has not been edited. A full set of peer reviewer comments for the entire history of the article can be found online at the link above. All code and data has been provided open source.

https://wellcomeopenresearch.org/articles/5-69

The paper in the chapter was written with a single co-author, Dr Paul Schneider. Robert Smith is the lead and corresponding author. Robert and Paul conceptualised the paper together in 2019, while working with the WHO on a project to improve the usability of an economic model of Female Genital Mutilation reduction strategies, while Robert was working on the HEAT walking and cycling model, and when working together with parkrun to optimise locations of events. Since the paper was more aligned to Robert's research, he led the process of developing the concept into a tutorial paper and obtaining input from external stakeholders via a workshop, one to one discussion with experts at UCL and Imperial College London, and submission and responses to reviewers in Wellcome Open Research and therefore was named as first author. Paul provided comments and suggestions throughout.

METHOD ARTICLE

REVISED Making health economic models Shiny: A tutorial [version 2; peer review: 2 approved]

Robert A. Smith*, Paul Schneider*

School of Health and Related Research, University of Sheffield, Regents Court, Sheffield, S1 4DA, UK

* Equal contributors

v2 **First published:** 14 Apr 2020, **5**:69
https://doi.org/10.12688/wellcomeopenres.15807.1
Latest published: 31 Jul 2020, **5**:69
https://doi.org/10.12688/wellcomeopenres.15807.2

Abstract

Health economic evaluation models have traditionally been built in Microsoft Excel, but more sophisticated tools are increasingly being used as model complexity and computational requirements increase. Of all the programming languages, R is most popular amongst health economists because it has a plethora of user created packages and is highly flexible. However, even with an integrated development environment such as R Studio, R lacks a simple point and click user interface and therefore requires some programming ability. This might make the switch from Microsoft Excel to R seem daunting, and it might make it difficult to directly communicate results with decisions makers and other stakeholders.

The R package Shiny has the potential to resolve this limitation. It allows programmers to embed health economic models developed in R into interactive web browser based user interfaces. Users can specify their own assumptions about model parameters and run different scenario analyses, which, in the case of regular a Markov model, can be computed within seconds. This paper provides a tutorial on how to wrap a health economic model built in R into a Shiny application. We use a four-state Markov model developed by the Decision Analysis in R for Technologies in Health (DARTH) group as a case-study to demonstrate main principles and basic functionality.

A more extensive tutorial, all code, and data are provided in a GitHub repository.

Keywords

Health Economics, R, RShiny, Decision Science

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 31 Jul 2020	✓ view	✓ view
version 1 14 Apr 2020	? view	✓ view

1. **Talitha L. Feenstra**, University of Groningen, Groningen, The Netherlands

2. **Yiqiao Xin**, University of Glasgow, Glasgow, UK

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert A. Smith (rasmith3@sheffield.ac.uk)

Author roles: Smith RA: Conceptualization, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Schneider P**: Conceptualization, Investigation, Methodology, Project Administration, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: R.S. and P.S. are joint funded by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science [108903] and the University of Sheffield.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Smith RA and Schneider P. **Making health economic models Shiny: A tutorial [version 2; peer review: 2 approved]** Wellcome Open Research 2020, **5**:69 https://doi.org/10.12688/wellcomeopenres.15807.2

First published: 14 Apr 2020, **5**:69 https://doi.org/10.12688/wellcomeopenres.15807.1

REVISED **Amendments from Version 1**

The changes made are all in response to the reviewers suggestions. For each comment by the reviewers we have responded with the exact change (see responses to reviewer comments below).

None of these changes change the fundamental teaching strategy of the tutorial, or the reasons why this tutorial paper was published, they are just relatively minor changes to the text used to clarify things which the reviewers found more difficult to follow, or felt we should have explained differently.

In two cases this resulted in changes to large chunks of text.

Running the model for a specific set of PSA inputs - We provide more detail on the way in which the f_MM_ sicksicker function works. This was in response to a request by reviewer 2 to do this.

Discussion – we add in more clarity about the limitations and strengths of the apps in response to the reviewers (in blue). This is in response to a request by reviewer 1 to do this.

Any further responses from the reviewers can be found at the end of the article

Introduction

As the complexity of health economic decision models increase, there is growing recognition of the advantages of using high level programming languages (e.g. R, Python, C++, Julia) to support statistical analysis. Depending on the model that is being used, Microsoft Excel can be relatively slow. Certain types of models (e.g. individual-level simulations) can take a very long time to run or become computationally infeasible, and some essential statistical methods can hardly be implemented at all (e.g. survival modelling, network meta-analysis, value of sample information), or rely on exporting results from other programs (e.g. R, STATA, WinBUGs).

Of all the high level programming languages, R is the most popular amongst health economists[1]. R is open source and supported by a large community of statisticians, data scientists and health economists. There are extensive collections of (mostly free) online resources, including packages, tutorials, courses, and guidelines. Chunks of code, model functions, and entire models are shared by numerous authors, which allow R users to quickly adopt and adapt methods and code created by others. Importantly for the UK, R is also currently the only programming environment accepted by NICE for HTA submissions, the alternative submission formats Excel, DATA, Treeage, and WinBUGs are all software applications[2].

Despite the many strengths of the script based approach (e.g R) to decision modelling, an important limitation has been the lack of an easy-to-understand user-interface, which would be useful as it "facilitates the development and communication of the model structure" (p.743)[1]. While it is common practice for 'spreadsheet models' to have a structured front tab, which allows decision makers to manipulate model assumptions and change parameters to assess their impact on the results, up until recently, R models had to be adapted within script files or command lines.

Released in 2012, Shiny is an R-package that can be used to create a graphical, web browser based interface. The result looks like a website, and allows users to interact with underlying R models without the need to manipulate the source code[3]. Shiny has already been widely adopted in many different areas and by various organisations to present the results of statistical analysis[4]. Within health economics Shiny is currently being used to conduct network meta analysis[5] and value of information analysis[6,7].

Using Shiny, it is possible to create flexible user interfaces that allow users to specify different assumptions, change parameters, run underlying R code and visualise results. The primary benefit of this is that it makes script based computer models accessible to those with no programming knowledge - opening models up to critical inquiry from decision makers and other stakeholders[8]. Other benefits come from leveraging the power of R's many publicly available packages; for example, allowing for publication quality graphs and tables to be downloaded, user specific data-files to be uploaded, open-access data to be automatically updated and, perhaps most importantly, to efficiently run comprehensive probabilistic sensitivity analyses in a fraction of the time that it would take in Microsoft Excel. Shiny web applications for R health economic decision models seem particularly useful in cases where model parameters are highly uncertain or unknown, and where analysis is conducted with heterogeneous assumptions (e.g. for different populations). Examples of well-designed shiny applications include, for example, the the Innovation and Value Initiative's open-source rheumatoid arthritis individual patient simulation model, Bresmed's 'IntRface' application, and the SHARP CKD-CVD outcomes model[9–11].

While, from a transparency perspective, it is preferable that models constructed in R are made open-access to improve replicability and collaboration, it is not a requirement[12]. Sensitive and proprietary data and/or models can

be shared internally, or through password-protected web applications, negating the need to email zipped folders. Once an R model and a Shiny application have been created, they can also be easily adapted, making it possible to quickly update the model when new information becomes available. Several authors have postulated that there is considerable potential in using Shiny to support and improve health economic decision making. Incerti *et al.* (2019) identified web applications as being an essential part of modelling, stating that they "believe that the future of cost-effectiveness modeling lies in web apps, in which graphical interfaces are used to run script-based models" (p.577)[13]. Similarly, Baio and Heath (2017) predicted that R Shiny web apps will be the "future of applied statistical modelling, particularly for cost-effectiveness analysis" (p.e5)[14]. Despite these optimistic prognoses, adoption of R in health economics has been slow and the use of Shiny seems to have been limited to only a few cases. A reason for this might be the lack of accessible tutorials tailored towards an economic modeller audience.

Here, we provide a simple example of a Shiny web app, using a general four-state Markov model. The model is based on the 'Sick-Sicker model', which has been described in detail in previous publications[15,16] and in open source teaching materials by the DARTH workgroup[17]. The model was slightly adapted to implement probabilistic sensitivity analysis. This paper aims to provide a tutorial, designed specifically for those familiar with decision modelling in R, to create web-based user interfaces for R models using R Shiny.

Methods

While the focus of this tutorial is on the application of Shiny for health economic models, below we provide a brief overview of the "Sick-Sicker model". For further details, readers are encouraged to consult previous publications by the DARTH group[15,16,18] and the DARTH group website[17].

The Sick-Sicker model is a four-state (Healthy, Sick, Sicker or Dead) time-independent Markov model. The cohort progresses through the model in cycles of equal duration, with the proportion of those in each health state in the next cycle being dependant on the proportion in each health state in the current cycle and a time constant transition probability matrix.

The analysis incorporates probabilistic sensitivity analysis (PSA) by creating a data-frame of PSA inputs (one row being one set of model inputs) based on cost, utility and state transition probability distributions using the function *f_gen_psa* and then running the model for each set of PSA inputs using the model function *f_MM_sicksicker*.We therefore begin by describing the two functions *f_gen_psa* and *f_MM_sicksicker* in more detail before moving on to demonstrate how to create a user-interface. In this tutorial, we follow Alarid-Escudero *et al.'s* (2019) coding framework and add to it the prefix 'f_' to denominate functions[15].

Functions

The *f_gen_psa* function (see the file f_gen_psa.R in the open access repository: https://doi.org/10.5281/zenodo.3727052[19]) returns a data-frame of probabilistic sensitivity analysis inputs: transition probabilities between health states using a beta distribution, hazard rates using a log-normal distribution, costs using a gamma distribution and utilities using a truncnormal distribution. It relies on two inputs, the number of simulations (PSA inputs), and the cost (which takes a fixed value). We set the defaults to 1000 and 50, respectively.

Running the model for a specific set of PSA inputs

The function *f_MM_sicksicker* (see the file f_MM_sicksicker in the open access repository: https://doi.org/10.5281/zenodo.3727052) makes use of the *with* function, which applies an expression (in this case the rest of the code) to a data-set (in this case params, which will be a row of PSA inputs).

The function first calculates transition probabilities from each health state to each health state and uses these to fill a transition probability matrix (*m_P*). It then creates a matrix for the markov trace (*m_TR*) which has t+1 nrows and four columns (one for each health state). The 'PROCESS' part of the code then 'loops' through the markov model, using matrix multiplication (elicited using %*% in R), iteratively computing, for each period, the proportion of the population that is in each state.

In the 'OUTPUT' section of the code the markov trace (*m_TR*) is multiplied (again using matrix multiplication) with vectors of health state utilities (e.g. v_u_trt) and costs (e.g. v_c_trt), giving a vector of total costs and utilities in each time interval. These vectors are then discounted using a discount weight vector (e.g. *v_dwe* & *v_dwe*) to arrive at a single cost/QALY value. The resulting total discounted costs and QALY estimates for the treatment and the no-treatment group then are combined into a vector and returned from the function. In this simple example,

treatment only influences utilities and costs, not transition probabilities. For further details on the underlying model, we refer to the published source code[19]

Creating the model wrapper

When using a web application, it is likely that the user will want to be able to change parameter inputs and rerun the model. In order to make this simple, we recommend wrapping the entire model into a function. We call this function *f_wrapper*, using the prefix *f_* to denote that this is a function.

The wrapper function has as its inputs all the parameters that we may wish to vary using R-Shiny. We set the default values to those of the base model in any report/publication. The model then generates PSA inputs using the *f_gen_psa* function, creates an empty table of results, and runs the model for each set of PSA inputs (a row from *df_psa*) in turn. The function then returns the results in the form of a data-frame with n=5 columns and n=psa rows. The columns contain the costs and QALYs for treatment and no treatment for each PSA run, as well as an ICER for that PSA run.

Model wrapper function

```
f_wrapper <- function(

#-- User adjustable inputs --#

# age at baseline
n_age_init = 25,
# maximum age of follow up
n_age_max  = 110,
# discount rate for costs and QALYS
d_r     = 0.035,
# number of simulations
n_sim   = 1000,
# cost of intervention treatment in states sick and sicker
c_Trt   = 50

 ){


#-- Unadjustable inputs --#

# number of cycles
n_t <- n_age_max - n_age_init
# the 4 health states of the model:
v_n <- c("H" , "S1" , "S2" , "D")
# number of health states
n_states <- length(v_n)

#-- Create PSA Inputs --#

df_psa <- f_gen_psa(n_sim = n_sim,
                    c_Trt =  c_Trt)

#-- Run PSA --#

# Initialize matrix of results outcomes
m_out <- matrix(NaN,
                nrow = n_sim,
                ncol = 5,
                dimnames = list(1:n_sim,
                c("Cost_NoTrt", "Cost_Trt",
                  "QALY_NoTrt", "QALY_Trt",
                  "ICER")))
```

```
# run model for each row of PSA inputs
for(i in 1:n_sim){

 # store results in row of results matrix
 m_out[i,] <- f_MM_sicksicker(df_psa[i, ])

  } # close model loop
#-- Return results --#

# convert matrix to dataframe (for plots)
 df_out <- as.data.frame(m_out)

# output the dataframe from the function
 return(df_out)

 } # end of function
```

Integrating into R-Shiny

The next step is to integrate the model function into a Shiny web-app. This is done within a single R file, which we call *app.R*. This can be found within the GitHub repository here.

The app.R script has three main parts, each are addressed in turn below:

- set-up (getting everything ready so the user-interface and server can be created)
- user interface (what people will see)
- server (R code running in the background)

Figure 1 depicts the relationship between the server and the user interface within the Shiny application. On a conceptual level, the user interface has three components: Shiny inputs (objects that the user can specify, e.g. by inputting a number), Shiny outputs (objects created on the server side, e.g. plots and tables), and non-interactive features (any fixed elements, such as texts, headings, logos etc.). The server works almost like a normal R session. It runs various R operations, including the model function, which takes non-Shiny inputs (defined only on the server side) and some Shiny inputs from the user interface. The results are then sent to the user interface and displayed as Shiny outputs.

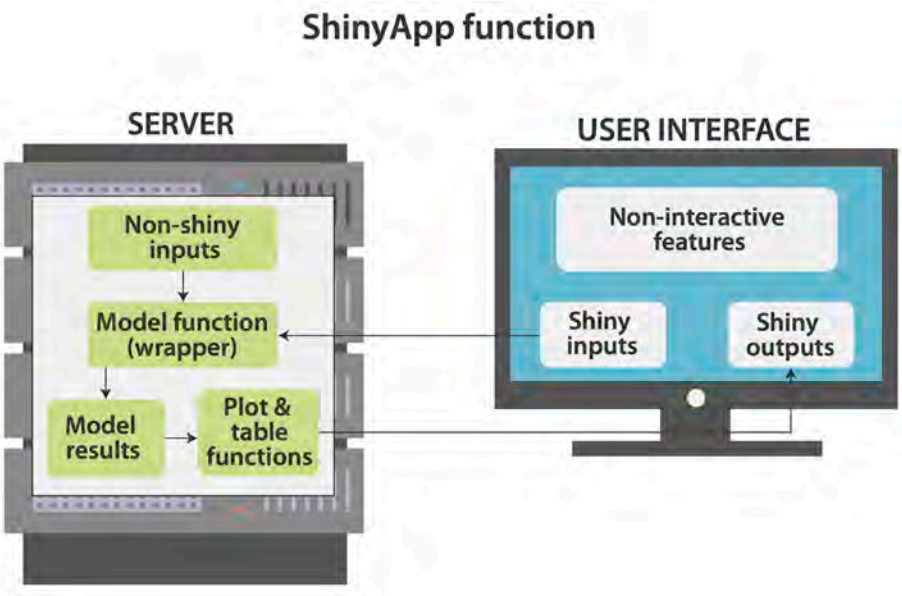


Figure 1. Diagram depicting how the Sick-Sicker app is structured.

Initial set-up

The set-up is relatively simple. First, load the R-Shiny package from your library so that you can use the *shinyApp* function. The next step is to use the *source* function in baseR to run the script that creates the *f_wrapper* function, being careful to ensure your relative path is correct ('./wrapper.R' should work if the wrapper.R file is in the same folder as the app.R file).

Code initialization (within app.R)

```
#  install 'shiny' if haven't already.
#  # install.packages("shiny")  # necessary if you don't already have the function
'shiny' installed.

# we need the function shiny installed, this loads it from the library.
library(shiny)

#  source the wrapper function.
  source("./wrapper.R")
```

Creating the user interface function

The user interface is extremely flexible, we show the code for a very simple structure (fluidpage) with a sidebar containing inputs and a main panel containing outputs. We have done very little formatting in order to minimize the quantity of code while maintaining basic functionality. In order to get an aesthetically pleasing application, we recommend much more sophisticated formatting, relying on CSS, HTML and Javascript.

The example user interface displayed in Figure 2 and online on this website. The user interface is a *fluidpage* in a *sidebarLayout* (other types of layout are available). The *sidebarLayout* is made up of two components, a titlepanel and a sidebar layout display (which itself is split into a sidebar and a main panel). This is a basic structure used for teaching purposes, there are a plethora of templates available online.

The title panel contains the title "Sick Sicker Model in Shiny", the sidebar panel contains two numeric inputs and a slider input ("Treatment Cost", "PSA runs", "Initial Age") and an action button ("Run / update model"). The values of the inputs have ID tags (names), which are recognised and used by the server function, we denote these with the prefix "SI" to indicate they are 'Shiny Input' objects (*SI_c_Trt*, *SI_n_sim*, *SI_n_age_init*). The action button also has an ID, this is not an input into the model wrapper *f_wrapper* so we leave out the SI and call it *run_model*.

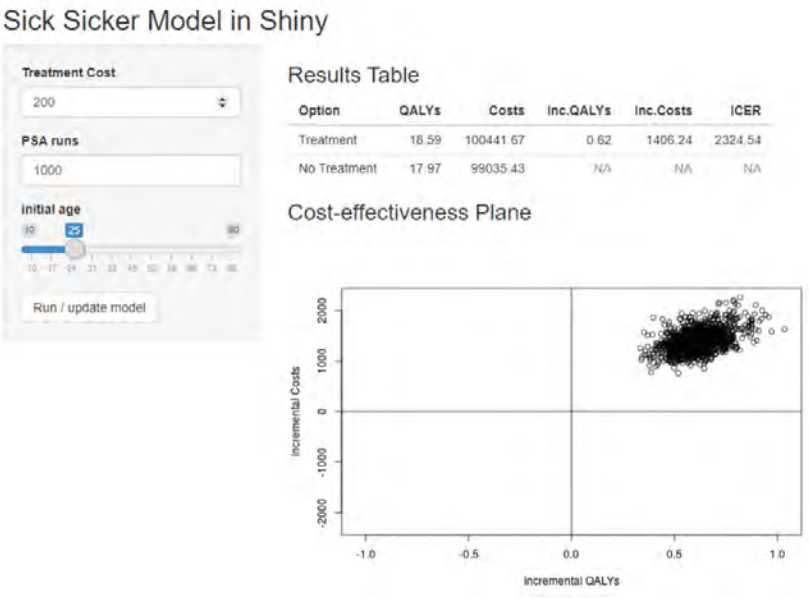


Figure 2. Screen-print of Sick-Sicker model user interface.

The main panel contains two objects that have been output from the server: *tableOutput("SO_icer_table")* is a table of results, and *plotOutput("SO_CE_plane")* is a cost-effectiveness plane plot. It is important that the format (e.g. *tableOutput*) matches the format of the object from the server (e.g. *SO_icer_table*). Again, the *SO* prefix reflects the fact that these are Shiny Outputs. The two *h3()* functions are simply headings, which appear as "Results Table" and "Cost-effectiveness Plane".

Shiny user interface code

```
ui <- fluidPage (    # creates empty page

  #  title of app
  titlePanel("Sick Sicker Model in Shiny"),

  # layout is a sidebar-layout
  sidebarLayout(

    sidebarPanel( # open sidebar panel

# input type numeric
    numericInput(inputId = "SI_c_Trt",
                 label = "Treatment Cost",
                 value = 200,
                 min = 0,
                 max = 400),

    numericInput(inputId = "SI_n_sim",
                 label = "PSA runs",
                 value = 1000,
                 min = 0,
                 max = 400),
# input type slider
    sliderInput(inputId = "SI_n_age_init",
                label = "Initial Age",
                value = 25,
                min = 10,
                max = 80),

# action button runs model when pressed
    actionButton(inputId = "run_model",
                 label   = "Run model")

                ),  # close sidebarPanel

# open main panel
  mainPanel(

# heading (results table)
    h3("Results Table"),

# tableOutput id = icer_table, from server
    tableOutput(outputId = "SO_icer_table"),

# heading (Cost effectiveness plane)
    h3("Cost-effectiveness Plane"),

# plotOutput id = SO_CE_plane, from server
    plotOutput(outputId = "SO_CE_plane")

            ) # close mainpanel

      ) # close side barlayout

  ) # close UI fluidpage
```

Creating the server function

The server is marginally more complicated than the user interface. It is created by a function with inputs and outputs. The observe event indicates that when the action button *run_model* is pressed the code within the curly brackets is run. The code will be re-run if the button is pressed again. Setting the parameter ignoreNULL to False lets the model run when it is initialised, i.e. when the app is started.

The first thing that happens when the *run_model* button is pressed is that the model wrapper function *f_wrapper* is run with the user interface inputs (*SI_c_Trt*, *SI_n_age_init*, *SI_n_sim*) as inputs to the function. The *input* prefix indicates that the objects have come from the user interface. The results of the model are stored as the data-frame object *df_model_res*.

The ICER table is then created and output (note the prefix *output*) in the object *SO_icer_table*. The function renderTable generates a table from the model results to display it on the web interface See previous section on the user interface and note that the *tableOutput* function has as an input *SO_icer_table*. The function *renderTable* rerenders the table continuously so that the table always reflects the values from the data-frame of results created above. In this simple example we have created a table of results using code within the script. Normally we would use a custom function that creates a publication quality table that is aesthetically pleasing. There are different packages that provide this functionality[15,20,21].

The cost-effectiveness plane is created in a similar process, using the *renderPlot* function to continuously update a plot, which is created using baseR plot function using incremental costs and QALYs calculated from the results dataframe *df_model_res*. For aesthetic purposes we recommend this is replaced by a ggplot2 or plotly plot, which have much improved functionality[22,23]. As with the results table, there are also numerous health economic modelling specific R packages that have plotting features[15,20,21].

Shiny server function

```
server <- function(input, output){

 # when action button pressed ...
  observeEvent(input$run_model,
               ignoreNULL = F, {

 # Run model function with Shiny inputs
  df_model_res = f_wrapper(
         c_Trt = input$SI_c_Trt,
         n_age_init = input$SI_n_age_init,
         n_sim = input$SI_n_sim)

 #-- CREATE COST EFFECTIVENESS TABLE --#

 # renderTable continuously updates table
  output$SO_icer_table <- renderTable({

  df_res_table <- data.frame( # create dataframe

  Option = c("Treatment","No Treatment"),

  QALYs  = c(mean(df_model_res$QALY_Trt),
             mean(df_model_res$QALY_NoTrt)),

  Costs  = c(mean(df_model_res$Cost_Trt),
             mean(df_model_res$Cost_NoTrt)),

  Inc.QALYs = c(mean(df_model_res$QALY_Trt) -
                mean(df_model_res$QALY_NoTrt),
                NA),
```

```
 Inc.Costs = c(mean(df_model_res$Cost_Trt) –
                mean(df_model_res$Cost_NoTrt),
                NA),

   ICER = c(mean(df_model_res$ICER), NA)

   ) # close data–frame
# round the data–frame to two digits
df_res_table[,2:6] = round(
        df_res_table[,2:6],digits = 2)

# print the results table
  df_res_table

  }) # table plot end.

#–– CREATE COST EFFECTIVENESS PLANE ––#

# render plot repeatedly updates.
output$SO_CE_plane <– renderPlot({

# calculate incremental costs and qalys
df_model_res$inc_C <– df_model_res$Cost_Trt –
                      df_model_res$Cost_NoTrt

df_model_res$inc_Q <– df_model_res$QALY_Trt –
                      df_model_res$QALY_NoTrt

# create cost effectiveness plane plot

plot(
# x y are incremental QALYs Costs
x = df_model_res$inc_Q,
y = df_model_res$inc_C,

# label axes
xlab = "Incremental QALYs",
ylab = "Incremental Costs",

# set x–limits and y–limits for plot.
xlim = c( min(df_model_res$inc_Q,
              df_model_res$inc_Q*–1),
          max(df_model_res$inc_Q,
              df_model_res$inc_Q*–1)),

ylim = c( min(df_model_res$inc_C,
              df_model_res$inc_C*–1),
          max(df_model_res$inc_C,
              df_model_res$inc_C*–1)),

# include y and y axis lines.
abline(h = 0,v = 0)
 ) # CE plot end

 }) # renderplot end

 }) # Observe event end


  } # Server end
```

Running the app

The app can be run within the R file using the function *shinyApp*, which depends on the *ui* and *server* that have been created and described above. Running this creates a Shiny application in the local environment (e.g. your desktop). It is also possible to deploy the application onto the web from RStudio using the shinyapps.io server (using the publish button in the top right corner of the R-file in R-Studio). Alternatively, apps can be hosted on private servers and integrated into existing websites. Server specifications should be chosen to match model requirements: while simple Markov chain state transition models may run on almost any server, more computationally burdensome models (e.g. agent-based models) may require considerable computing power. A step by step guide to the process of publishing applications can be found on the R-Shiny website or other online resources[3,24].

Running the app

```
shinyApp(ui , server)
```

Additional functionality

The example Sick-Sicker web-app that has been created is a simple, but functional, R-Shiny user interface for a health economic model. There are a number of additional functionalities, many of which are covered in an online book by Hadley Wickham[24].

- fully customised user interface aesthetics. Since the user interface is translated into HTML and CSS it is possible to customise all components (such as colors, fonts, graphics, layouts and backgrounds)[3,25].
- leverage many popular R packages to visualise model inputs (e.g. distributions) and outputs (e.g. plots and results tables)[22,23,26].
- upload files containing input parameters and data to the app[24].
- download specific figures and tables from the app[24].
- create a downloadable full report including model inputs and outputs[24].
- send model results/report to an email address once the model has finished running[27].

It is also possible to integrate all of the steps of health economic evaluation into one program. After selecting a subgroup of studies to use as inputs for a network meta-analysis, and economic model assumptions, the user would be required to simply click a 'run' button. They would then be presented with results of the network meta-analysis, economic model and value of information analysis in one simple user-interface. The app user would then also be able to download a report (or have it sent to an email address) with the model results and appropriate visualisations updated to reflect their assumptions.

Discussion

In this paper, we demonstrated how to generate a user-friendly interface for an economic model programmed in R, using the Shiny package. This tutorial shows that the process is relatively simple and requires limited additional programming knowledge than that required to build a decision model in R.

The movement towards script based health economic models with web based user interfaces is particularly useful in situations where a general model structure has been created with a variety of stakeholders in mind, each of which may have different assumptions (input parameters) and wish to conduct sensitivity analysis specific to their decision. For example, the World Health Organisation Department of Sexual and Reproductive Health and Research recently embedded a Shiny application into their website[28]. The application runs a *heemod* model[20] in R in an external server, and allows users to select their country and specify country specific assumptions (input parameters), run the model and display results.

A well designed user interface can allow users to explore and better understand the relationship between model input and results. This allows users to tailor the health economic model to their specific situation and assumptions, without the expense of creating a new model. This may be particularly useful in the following scenarios: Firstly, in areas where one health economic decision model is applied in range of circumstances (e.g. in public health, models are often built to be used in a number of different countries). Secondly, when the full model source code can or may not be shared (e.g. for proprietary or privacy reasons). R-Shiny apps can be made available in a way that allows users to interact with the web interface, without revealing the model behind it. Finally, R shiny apps may enable stakeholders and decision makers, who would otherwise not be able to interact directly with statistical computer models, to experiment with and to reflect on various scenarios and the validity of model inputs and outputs.

However, in all of these scenarios, the ability for users to test different assumptions is not without limits: the available options to vary point estimates or the uncertainty around input parameters are defined by the model developer, and it is also not possible to specify alternative model structures or test any other aspect that the developer did not implement. Therefore, the model source code and data will still need to be made available to reviewers to allow for a thorough assessments of health economic models. Further investigation into how to communicate economic decisions models in a transparent and inclusive way is an important avenue of future research

The authors' experience of creating user-interfaces for decision models has led to the conclusion that the most efficient method is to work iteratively, starting with a very simple working application, and adding functionality step by step, testing the app at each iteration to ensure it works as intended. It is worth noting that the simple model chosen as an exemplar is a markov model, however the method described can be applied to any model built using R, regardless of model type. For example, in this case, the *f_MM_sicksicker* function could also be replaced by a function containing any other type of model (e.g. a DES model).

There are several challenges that exist with the movement toward script based models with web-based user-interfaces. The first is the challenge of up-skilling health economic modeller used to working in Microsoft Excel. We hope that this tutorial provides a useful addition to previous tutorials demonstrating how to construct decision models in R[16]. A second, and crucial challenge to overcome, is a concern about deploying highly sensitive data and methods to an external server. While server providers such as ShinyIO provide assurances of SSR encryption and user authentication clients with particularly sensitive data may still have concerns. This problem can be avoided in two ways: firstly, if clients have their own server and the ability to deploy applications they can maintain control of all data and code, and secondly, the application could simply not be deployed, and instead simply created during a meeting using code and data shared in a zip file. Finally, a challenge (and opportunity) exists to create user-interfaces that are most user-friendly for decision makers in this field; this is an area of important research that requires closer collaboration between decision makers, stakeholders and health economic decision model developers.

Conclusion

The creation of web application user interfaces for health economic models constructed in high level programming languages should improve their usability, allowing stakeholders and third parties with no programming knowledge to conduct their own sensitivity analysis remotely. This tutorial provides a reference for those attempting to create a user interface for a health economic decision model created in R. Further work is necessary to better understand how to design interfaces that best meet the needs of different decision makers.

Data availability

All data underlying the results are available as part of the article and no additional source data are required.

Software availability

Source code available from: https://github.com/ RobertASmith/paper_makeHEshiny

Archived source code at time of publication: https://doi.org/10.5281/zenodo.3730897[19].

License: MIT license

Acknowledgements

We would like to thank Gianluca Baio, Nathan Green and all participants in the pilot tutorial held at the University of Sheffield for comments. All errors are the responsibility of the authors.

References

1. Jalal H, Pechlivanoglou P, Krijkamp E, *et al.*: **An Overview of R in Health Decision Sciences.** *Med Decis Making.* 2017; **37**(7): 735–746.
PubMed Abstract | Publisher Full Text

2. National Institute for Health and Care Excellence (Great Britain): **Guide to the processes of technology Appraisal [Internet]**. *National Institute for Health and Care Excellence*, 2014.
PubMed Abstract

3. Beeley C: **Web application development with R using Shiny.** *Packt Publishing Ltd.* 2016.
Reference Source

4. Gendron J: **Introduction to R for Business Intelligence.** *Packt Publishing Ltd.* 2016.
Reference Source

5. Owen RK, Bradbury N, Xin Y, *et al.*: **Metainsight: An interactive web-based tool for analyzing, interrogating, and visualizing network meta-analyses using r-shiny and netmeta.** *Res Synth Methods.* 2019; **10**(4): 569–581.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Strong M, Oakley JE, Brennan A: **Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach.** *Med Decis Making.* 2014; **34**(3): 311–326.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Baio G, Berardi A, Heath A: **Bceaweb: A user-friendly web-app to use bcea.** In: *Bayesian Cost-Effectiveness Analysis with the R package BCEA*, Springer, 2017; 153–166.
Publisher Full Text

8. Jansen JP, Incerti D, Linthicum MT: **Developing Open-Source Models for the US Health System: Practical Experiences and Challenges to Date with the Open-Source Value Project.** *Pharmacoeconomics.* 2019; **37**(11): 1313–1320.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Hart R, Burns D, Ramaekers B, *et al.*: **R and Shiny for Cost-Effectiveness Analyses: Why and When? A Hypothetical Case Study.** *PharmacoEconomics.* 2020; **38**(7): 765–776.
PubMed Abstract | Publisher Full Text

10. Incerti D, Curtis JR, Shafrin J, *et al.*: **A flexible open-source decision model for value assessment of biologic treatment for rheumatoid arthritis.** *Pharmacoeconomics.* 2019; **37**(6): 829–843.
PubMed Abstract | Publisher Full Text

11. Schlackow I, Kent S, Herrington W, *et al.*: **A policy model of cardiovascular disease in moderate-to-advanced chronic kidney disease.** *Heart.* 2017; **103**(23): 1880–1890.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Hatswell AJ, Chandler F: **Sharing is caring: the case for company-level collaboration in pharmacoeconomic modelling.** *Pharmacoeconomics.* 2017; **35**(8): 755–757.
PubMed Abstract | Publisher Full Text

13. Incerti D, Thom H, Baio G, *et al.*: **R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment.** *Value Health.* 2019; **22**(5): 575–579.
PubMed Abstract | Publisher Full Text

14. Baio G, Heath A: **When simple becomes complicated: why excel should lose its place at the top table**. 2017.
Publisher Full Text

15. Alarid-Escudero F, Krijkamp EM, Pechlivanoglou P, *et al.*: **A need for change! a coding framework for improving transparency in decision modeling.** *Pharmacoeconomics.* 2019; **37**(11): 1329–1339.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Alarid-Escudero F, Krijkamp EM, Enns EA, *et al.*: **Cohort state-transition models in R: From conceptualization to implementation.** *arXiv preprint arXiv: 2001.* 2020; 07824.
Reference Source

17. DARTH Workgroup: **Decision analysis in r for technologies in health.** 2020.
Reference Source

18. Krijkamp EM, Alarid-Escudero F, Enns EA, *et al.*: **Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial.** *Med Decis Making.* 2018; **38**(3): 400–422.
PubMed Abstract | Publisher Full Text | Free Full Text

19. Robert S: **RobertASmith/paper_makeHEshiny: Making health economics Shiny: a tutorial**. 2020.
http://www.doi.org/10.5281/zenodo.3730897

20. Filipovic-Pierucci A, Zarca K, Durand-Zaleski I: **Markov models for health economic evaluation modelling in r with the heemod package.** *Value Health.* 2016; **19**(7): A369.
Publisher Full Text

21. Baio G, Berardi A, Heath A: **Bayesian Cost-Effectiveness Analysis with the R package BCEA.** Springer, New York, NY, 2017; ISBN 978-3-319-55718-2.
Publisher Full Text

22. Sievert C: **plotly for R**. 2018.
Reference Source

23. Wickham H: **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016; ISBN 978-3-319-24277-4.
Reference Source

24. Wickham H: **Mastering shiny.** 2020.
Reference Source

25. Shiny: **Build your entire ui with html.** 2020.
Reference Source

26. Dowle M, Srinivasan A: **data.table: Extension of 'data.frame'.** 2019. R package version 1.12.8.
Reference Source

27. Hester J: **gmailr: Access the 'Gmail' 'RESTful' API.** 2019. R package version 1.0.0.
Reference Source

28. World Health Organization: **FGM Cost Calculator**. 2020; [Accessed 22 July 2020].
Reference Source

The Sick Sicker model used in this tutorial is a simple open-source model that has been widely used for teaching health economists to use R (Alarid-Escudero et al. 2023). The goal of using the Sick Sicker Model was to make the paper more accessible to readers. The approach can also be applied to more complex models and could be used to create a user-interface for the PACEM model discussed previously. However, since the PACEM model was not designed to inform commissioning decisions and does not have a decision maker to aid, a user-interface has not been developed. In the future, if the PACEM model is used to inform policy, a new user-interface could be created to allow users to conduct their own sensitivity analysis or upload data specific to their own region's population and physical activity levels.

Chapter 8. A tutorial on automating updates for health economic models while protecting sensitive data.

One of the main barriers to the transparency of health economic modelling is the need to protect sensitive data. Health economic models built for health economic evaluation, for example those submitted to agencies such as NICE, are often created using spreadsheet software (Jalal et al. 2017), which makes it difficult to separate the data from the model (Incerti et al. 2019). On the other hand, script-based models can be more easily separated from the underlying data they are based on, since the model code and the data are two separate entities. This has two main advantages: first, the model code can be shared without revealing sensitive data, improving the transparency and accessibility of health economic models (Cohen et al. 2017; Sampson et al. 2019); and second, script-based models can be automated to re-run every time new information becomes available, reducing the financial and administrative burden of model updates. While academics and experts in health economics may be particularly interested in the first advantage, agencies and academic groups and consultants struggling with workloads may be more excited about the second advantage.

In the context of economic evaluations of physical activity interventions, developing this methodology may help keep models up to date as new information emerges on the cost of treatment of different health conditions, baseline epidemiological parameters and as new evidence emerges on the relationship between physical activity and health outcomes. It may also be useful where evaluations occur in real time to decide whether or not to extend funding for a particular pilot study, for example providing free passes to a gym. By combining the health economic model into the trial stage of the project it may be possible to immediately determine whether to continue funding rather than delaying a decision to await results of a modelling study.

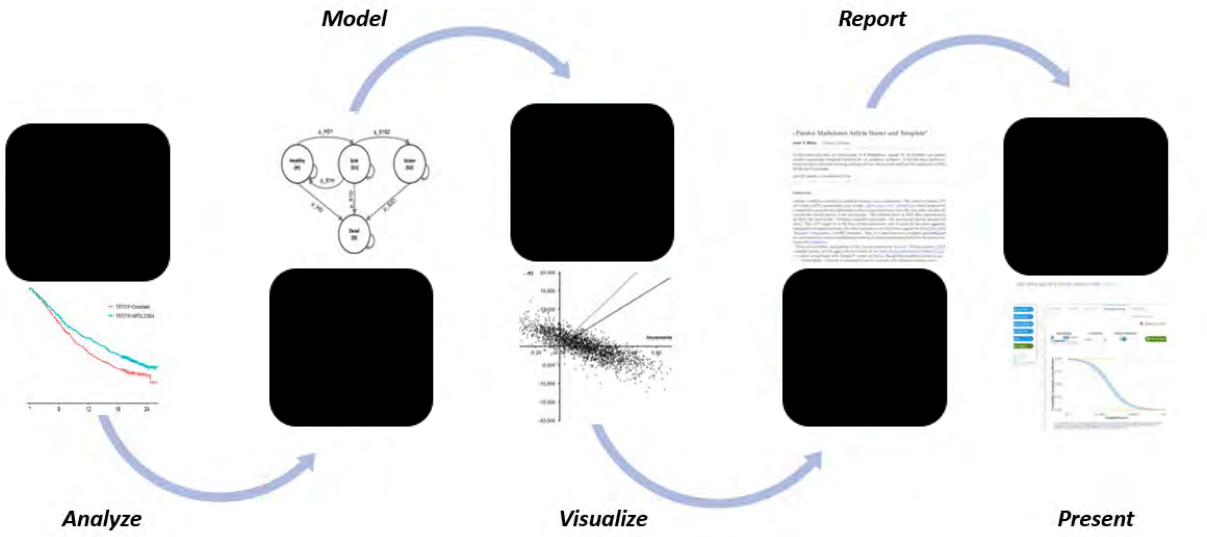
In this chapter, I describe a process for generating automated health economic evaluation reports when triggered by human interaction with a web app or changes to data stored on a remote server, such as in a pharmaceutical company's database. The method builds upon the development of web-based user interfaces for hosting health economic models (Smith & Schneider, 2020a) and applies the motivation behind the OpenSafely initiative (Williamson et al. 2020), which allows for the analysis of electronic health records data, to health economics and decision science. A paper, which includes all source code and data, has been published open access in Wellcome Open Research (Smith, Schneider & Mohammed, 2022a).

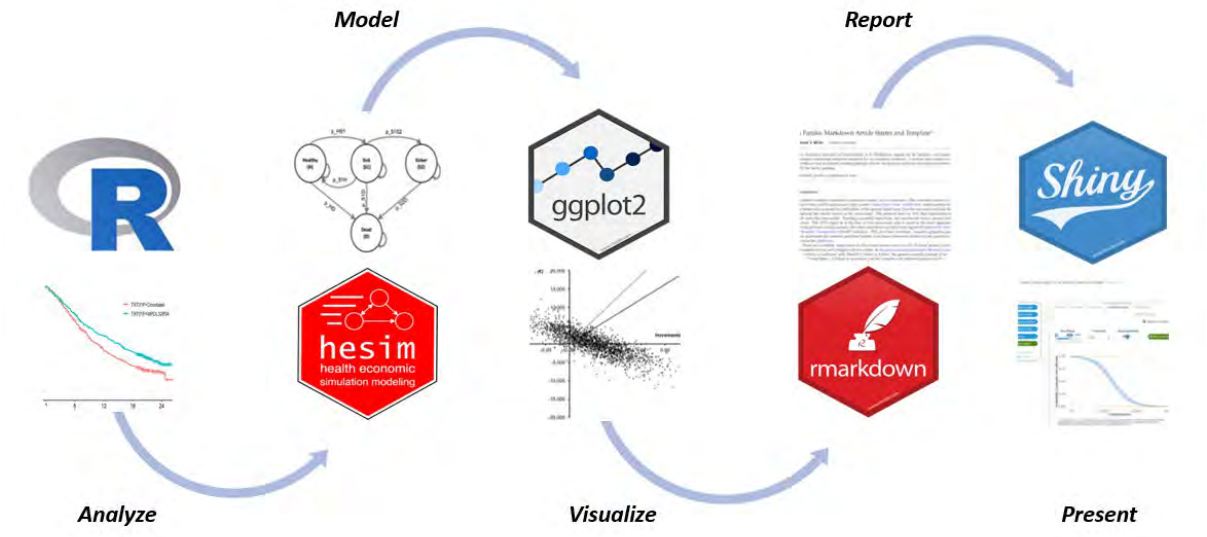
The goal of this process is to eliminate the need for companies to repeatedly share sensitive data with those building, reviewing, and using health economic evaluation models. It is hoped that those who own sensitive data will be more willing to allow stakeholders to interact with health economic models (e.g.

through a web app) or share their underlying code if they have confidence that the sensitive data remains secure. In the long term, this may lead to improved modelling transparency and a better understanding of the sensitivity of results to parameter values (Sampson et al., 2019). In addition, the process allows for the comparison or combination of multiple models using weights informed by goodness of fit to observed data or expert judgement (Jackson et al. 2009), helping to improve understanding of structural uncertainties as undertaken in processes like the Mount Hood (Kent et al. 2019).

Another advantage of this process is that it creates automated updates to health economic evaluation reports. The model can be re-run at a set time or when triggered by human interaction or data updates, and an updated model report generated. This reduces the time and resources required to update models as real-world evidence, clinical trial data, or parameter assumptions change. By semi-automating model updates, moving from manual copy-paste of results across a pipeline to a semi-automated update process as shown in Figure 14 below, health economic evaluation can become more efficient, allowing for more regular evaluation of new technologies with the same human resources, thereby avoiding health economic evaluations being ‘out-of-date’ (Shields, et al. 2022) and thereby improving external-validity. The effect of this should be to improve decisions by government medicines agencies and public health commissioners, ultimately leading to better population health outcomes.

Figure 14 The top visual depicts the commonly used process of conducting statistical analysis in STATA or SAS, but then building the underlying model in MS Excel, perhaps using VBA, creating plots and tables in MS Excel, and manually copy-pasting them into MS Word, and later presenting using MS PowerPoint. The bottom visual illustrates a process where every stage is run in the R software environment using open-source packages, enabling updates to model reports, presentations, and web applications to be automated or updated with just one click.





The concept of a "living" evidence base, which continually updates as new data becomes available, gained popularity during the COVID-19 pandemic as the evidence base rapidly expanded and decisions had to be made quickly (Eldridge et al. 2022). A recent paper I published with researchers from ScHARR, ICER, and NICE outlined a framework for a full "living" HTA pipeline and identified potential challenges and barriers to its adoption (Thokala et al. 2023). The paper discussed how each stage of the HTA pipeline could be made "living" and how updates could be semi-automated as new evidence becomes available. As far as the authors know, there are no published methods for automating any stage of the HTA pipeline. However, "living" systematic reviews are becoming more common (Eldridge et al. 2022) and there is a growing codebase that could be used to create "living" meta-analysis *(*Siemieniuk et al. 2020*)*. With the addition of Smith, Schneider & Mohammed's (2022a) contribution on a "living" health economic evaluation, it may not be long before the entire HTA process is combined into a single "living" (and possibly semi-automated) framework.

Living HTA: Automating Health Economic Evaluation with R

This section of the chapter contains a publication which outlines a method by which updates to health economic evaluation models can be automated without requiring sharing of data between data-owners and health economists, thus improving the efficiency and potentially transparency of health economic modelling.

This article was published open access following the requirement of the Wellcome Trust who financially supported this work. The conditions of the open access publishing allow use of the final published PDF, original submission, or accepted manuscript in this thesis (including in any electronic institutional repository or database). The original content has not been edited. A full set of peer reviewer comments for the entire history of the article can be found online at the link above. All code and data has been provided open source.

https://wellcomeopenresearch.org/articles/7-194

The paper was written with two co-authors: Paul Schneider and Wael Mohammed. Robert Smith was lead author and corresponding author and led the conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualisation, wrote the original draft and edited the manuscript based upon comments from other authors and reviewers. Paul Schneider contributed to the conceptualisation, methodology, software, reviewing the initial draft and reviewing the manuscript. Wael Mohammed contributed to the methodology, reviewing the initial draft and the responses to reviewers.

METHOD ARTICLE

REVISED Living HTA: Automating Health Economic Evaluation with R [version 2; peer review: 2 approved]

Robert A. Smith [1-3], Paul P. Schneider [1,3], Wael Mohammed [1,3]

[1]School of Health and Related Research, University of Sheffield, Sheffield, S1 4DA, UK
[2]Lumanity, Sheffield, S1 2GQ, UK
[3]Dark Peak Analytics, Sheffield, S11 7BA, UK

v2 **First published:** 21 Jul 2022, **7**:194
https://doi.org/10.12688/wellcomeopenres.17933.1
Latest published: 11 Oct 2022, **7**:194
https://doi.org/10.12688/wellcomeopenres.17933.2

Abstract

Background: Requiring access to sensitive data can be a significant obstacle for the development of health models in the Health Economics & Outcomes Research (HEOR) setting. We demonstrate how health economic evaluation can be conducted with minimal transfer of data between parties, while automating reporting as new information becomes available.
Methods: We developed an automated analysis and reporting pipeline for health economic modelling and made the source code openly available on a GitHub repository. The pipeline consists of three parts: An economic model is constructed by the *consultant* using pseudo data. On the data-owner side, an application programming interface (API) is hosted on a server. This API hosts all sensitive data, so that data does not have to be provided to the *consultant*. An automated workflow is created, which calls the API, retrieves results, and generates a report.
Results: The application of modern data science tools and practices allows analyses of data without the need for direct access – negating the need to send sensitive data. In addition, the entire workflow can be largely automated: the analysis can be scheduled to run at defined time points (e.g. monthly), or when triggered by an event (e.g. an update to the underlying data or model code); results can be generated automatically and then be exported into a report. Documents no longer need to be revised manually.
Conclusions: This example demonstrates that it is possible, within a HEOR setting, to separate the health economic model from the data, and automate the main steps of the analysis pipeline.

Keywords

HEOR, HTA, APIs, R, plumber

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 11 Oct 2022	✓ view	
version 1 21 Jul 2022	? view	✓ view

1. **Mohsen Sadatsafavi**, University of British Columbia, Vancouver, Canada

2. **Devin Incerti**, EntityRisk Inc., Princeton, USA

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Robert A. Smith (rasmith3@sheffield.ac.uk)

Author roles: Smith RA: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Schneider PP**: Conceptualization, Methodology, Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Mohammed W**: Methodology, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: R.A.S. Is part of the Scientific Committee for R for HTA, an academic consortium whose main objective is to explore the use of R for cost-effectiveness analysis. P.P.S. and W.M have no competing interests to declare.

Grant information: This work was jointly supported by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science (PHEDS) [108903, https://doi.org/10.35802/108903; 224853] and the University of Sheffield.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Smith RA, Schneider PP and Mohammed W. **Living HTA: Automating Health Economic Evaluation with R [version 2; peer review: 2 approved]** Wellcome Open Research 2022, **7**:194 https://doi.org/10.12688/wellcomeopenres.17933.2

First published: 21 Jul 2022, **7**:194 https://doi.org/10.12688/wellcomeopenres.17933.1

REVISED **Amendments from Version 1**

Since the previous version of the article we have done the following:

- Added additional text in the introduction section to outline the value of the contribution.
- Included a new figure, related to the above, which is more aesthetically pleasing than the one in the preprint.
- Refined the methods sections to improve the flow and clarity of the article, in particular we have added a paragraph to provide more information about the example use case for which we provide open source code.
- Added descriptions to the code chunks to make it clearer who is running the code chunks and where.
- Indented several lines of the code to make the code easier to follow.
- Added to the discussion section to include more information on deploying these types of APIs, and the limitations of different deployment methods.

The fundamental contribution of the paper, the general message and the open source code remains unchanged.

Any further responses from the reviewers can be found at the end of the article

Introduction

The development of economic models sometimes involves the transfer of sensitive data (e.g. individual patient or price data) between parties. This paper demonstrates how the use of application programming interfaces (API) allows data-owners in the Health Economics & Outcomes Research (HEOR) industry to collaborate with multiple partners on health economic decision models, while, retaining full control of their data. The use of an API furthermore makes it possible to streamline and automate reporting as new information becomes available, significantly reducing the financial and administrative burden of economic model updates.

To our knowledge this is the first publication to outline a process for automated reporting in HEOR, which we term Living HTA, and the first to demonstrate the process of sending health economic model algorithms to sensitive data using APIs.

Two other bodies of work are particularly relevant. The first is the OpenSafely initiative, which inspired this work. Williamson *et al.*[1] describe the OpenSafely interface, which was developed to analyse electronic health records data without the need to share confidential patient information:

> "secure software interface that allows detailed pseudonymized primary care patient records to be analysed in near-real time where they already reside - hosted within the highly secure data centre of the electronic health records vendor — to minimize the reidentification risks when data are transported off-site".

The method described in this paper has a similar objective, but aims to protect sensitive information in the HEOR sector.

The second work, a publication by Adibi *et al.*[2], describes a cloud-based model accessibility platform for models developed in R. The authors make the case for cloud based platforms to improve the accessibility, transparency and standardization of health economic models, particularly highlighting the benefits of hosting computationally burdensome models on remote servers. The authors outline a framework for hosting models, contained within R packages, which are run using calls to an API. A set of standardized model call functions provide the user of the API with enough information to pass the necessary parameters to the model, run the model, and retrieve the necessary results directly into an R session. The publication is the first, to our knowledge, to discuss the enormous implications that remote model hosting could have in the HEOR industry.

We combine elements from both Adibi *et al.*[2] and the OpenSafely initiative, and provide an open-source code base which demonstrates the ease with which APIs can be deployed on remote servers to avoid the need to share sensitive data, and enabling automation of model updates. In short, we propose that *data-owners* (e.g. pharmaceutical companies or governments), with support from health economists, host their own model accessibility platforms. We therefore see the primary contribution of this paper as being the development of a system in which the health economic model and the data are two separate entities, and the health economic model is sent to the data rather than the other way around. By working in this way, it is theoretically easier to share the model without the sensitive data on which it is run, although making the model open source is not a requirement. Our hope is that providing these materials will encourage others to use these methods to improve the transparency, accessibility and efficiency of health economic models.

Methods

A common problem in health technology assessment is a situation in which a data-owner (e.g. a company in the pharmaceutical industry) holds sensitive data, but requires the services of a consultant to conduct health economic modeling. For example the data-owner may have interim clinical trial data at the patient level, and may need an external health economist (the consultant) to build a state transition model to determine the cost-effectiveness of the treatment. Currently the data-owner may be required to send the consultant cuts of data as the trial progresses (e.g. 12 month, 18 month, 24 month). This is burdensome and results in multiple iterations of sharing sensitive data. We propose a solution that does not require the sharing of data between parties and allows for automated updates to the analysis as data is updated.

This automated analysis and reporting pipeline for health economic modeling consists of three parts:

- An economic model. The model can initially be developed using pseudo data – that is, randomly generated data, which has the same format as the actual data, but does not contain any sensitive information.
- An API, hosted by the data-owner side. It can be generated using the R package plumber. An automated workflow is created. This workflow sends the economic model to the data-owner's API. The model is then run within the data-owner's server. The results are sent back to the *consultant*, and a (PDF) report is automatically generated using RMarkdown[3]. This API server hosts all sensitive data, so that data does not have to be sent between parties.
- All of these processes can be controlled with a web-based user-interface. We provide an example user-interface built in the R shiny package[4], based on the tutorial application in our previous paper[5]. This application allows users to select input parameters with which to query the API, and view the results. This allows non-technical stakeholders to interact with the model in real time, while allowing the data-owner to retain control of the data. The application will always reflect the data on the data-owner's server, and the model hosted by the *consultant* at the time of use.

Figure 1 shows a schematic of the interaction between the data-owner's API and the *consultant's* automated workflow. All of the methods discussed in this paper, as well as the code for the demonstration app can be found contained within an open access GitHub repository (see *Software availability*[6]).

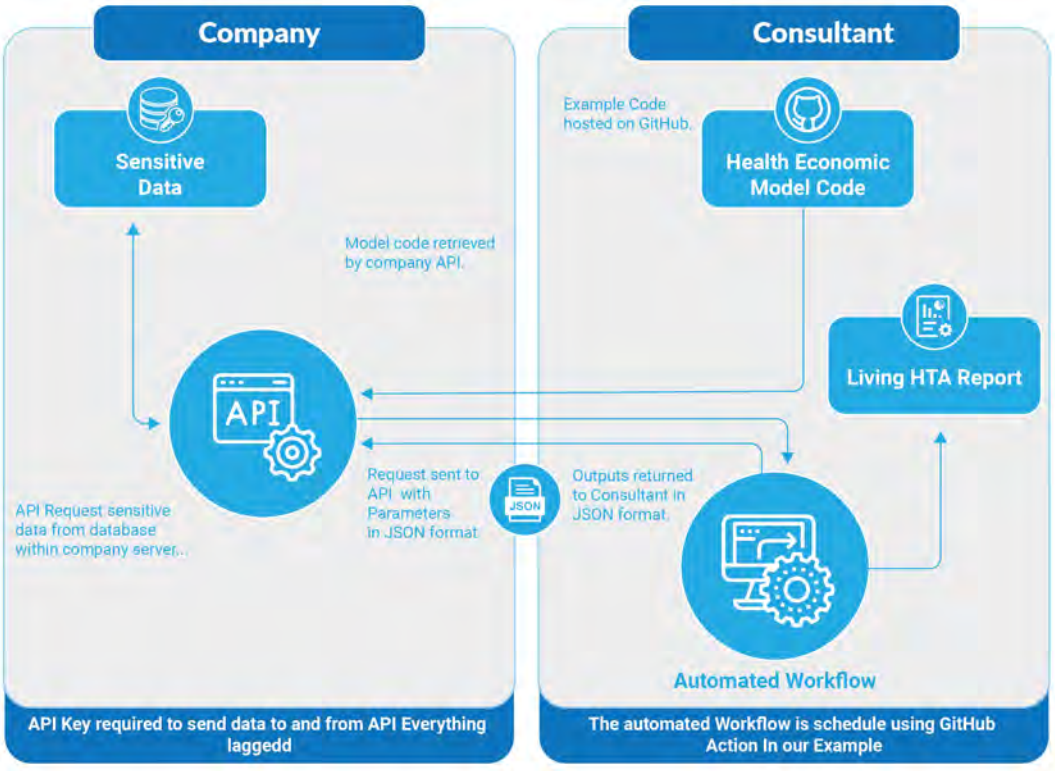


Figure 1. Schematic showing the interaction between the company API (application programming interface) and the consultant automated workflow. HTA, Health Technology Assessment; JSON, Javascript Object Notation.

The economic model

This model code has been adapted from the Decision Analysis in R for Technologies in Health (DARTH) group's open source Cohort state-transition model (the Sick-Sicker Model) which is discussed in Alarid-Escudero *et al.*[7] with open source code available online[8]. The code includes several functions, but for the purpose of this example we can treat the model as a black box, as a single function called *run_model* which runs the DARTH Sick Sicker model. The *run_model* function takes a single argument, *psa_inputs*, which is a data-frame containing Probabilistic Sensitivity Analysis parameter inputs for the model variables that are allowed to vary. Additional, sensitive parameters (including treatment costs and hazard ratio for treatment B) are not allowed to be varied by the API request and will be informed by the values held by the data-owner.

The data-frame has four columns:

- *parameter* - the name of the parameter (e.g. p_HS1)
- *distribution* - the distribution of that parameter (e.g. "beta")
- *V1* - the first parameter for the distribution in R (for beta this would be *shape1*, for normal this would be *mean*)
- *V2* - the second parameter for the distribution in R (for beta this would be *shape2*, for normal this would be *sd*)

The *run_model* function returns a data-frame with six columns. The first three columns are costs for each treatment option, and the second three columns are Quality Adjusted Life Years (QALYs) for each treatment option. Each row represents the result of the model run for a set of inputs.

The function described is designed as a simple reproducible example. The proposed method is flexible to any inputs, model structure, and outputs.

The API

An application programming interface is a set of rules, in the form of code, that allow different computers to interact with one another in real time. Whereas user-interfaces such as those generated by the R package *shiny* allow humans to interact with data, APIs are designed to enable computers to interact with data[4].

When a 'client' application wants to access data, it initiates an API call (*request*) via a web-server, to retrieve the data. If this request is deemed valid, the API makes a call to an external program/server, the server sends a response to the API with the data, and the API transfers the data to the 'client' application. In a sense, the API is the broker (or middle-man) between two systems.

There are numerous benefits to APIs:

- in supporting programmatic access. In contrast to what web applications offer (for example shiny apps), APIs allow users to access data, or other utilities (for example, proprietary applications) programmatically. Programmatic access enables users to invoke actions through an application or third-party tool. For example, R users can write a function that fetches or analyses data via an API and use it in their workflow as any other user-defined function.
- in allowing cross-platform communications. Statisticians and decision-model developers can use different programming languages or packages. For example, APIs can allow a decision analytic model, developed in C++ to programmatically utilise data from a bayesian meta-analysis performed using the Python programming language.

- in aiding speed of collaboration between institutions, ensuring inputs and outputs are standardised so that applications can 'talk' to one another. Users from one institution need not to take into account the software or package used by their partners, but focus on how they would interact with the expected data.
- in security, eliminating the necessity to share data manually (e.g. via email). All interaction with data can be logged and access can be restricted by passwords and by limiting IP address access. For example, APIs can safely allow statisticians to programmatically accumulate sub-group summary-statistics from securely stored trial-data to inform a network meta-analysis.
- in expanding sharing avenues. For example, APIs can allow institutions to give limited access to their proprietary tools such as in-house decision-analytic models. Users of such tools can pass their data to the model and receive the respective outputs via the API.
- eliminating computational burden on the client side (since all computation is done on the API owner side).

There are lots of different implementations of APIs, but the main focus of this paper is on *Partner APIs*, which are created to allow data transfer between two different institutions. This requires a medium level of security, usually through the creation of access keys that are shared with partners.

In the examples below we use Javascript Object Notation (JSON), a data interchange format that is commonly used to transfer information between computers, to pass information to and from our API. Since the model is written in R, we convert back and forth between JSON and R data formats using the *jsonlite* R package[9].

Creating the API using plumber

The R package *plumber* allows programmers to create web APIs by decorating R source code with roxygen-like comments[10,11]. These functions are then made available as API endpoints by plumber.

The API can be called using a number of HTTP request methods (also known as HTTP verbs). The most-commonly used methods POST, GET, PUT, PATCH, and DELETE correspond to create (POST and PUT), read (GET), update (PATCH), and delete (DELETE) operations. These annotations generate the API's endpoint(s) and specify the operation(s) or response(s) the respective R function is responsible for generating. The below example shows the 'GET' request (the default for web-browsers).

The code below gives an example function which echos a message. The function takes one input, a string with the message, and outputs the message contained within a list. If this function was created in R it would return a list containing some text, like this: The message is: 'example_msg'.

```
#* Echo back the input
#* @param msg The message to echo
#* @get /echo
function(msg="") {
  list(msg = paste0("The message is: '", msg, "'"))
}
```

The code for the model function uses the same principles, but is much more developed. There are three arguments to the model API; *path_to_psa_inputs*, *model_functions* and *param_updates*.

The core API function created by plumber sources the model functions from software development website GitHub, obtains the model parameter data from within the API, and then overwrites the rows of the parameter updates that exist in *param_updates*. It then runs the model functions using the updated parameters, post-processes the results, checks that no sensitive data is included in the results, and then returns a data-frame of results. This entire process occurs in the server on which the API is hosted, with inputs and outputs passed to the API over the web in JSON format.

Code chunk 1 - Generating the API **(this code is run on the data-owner's server)**

```
library(dampack)
library(readr)
library(assertthat)

#* @apiTitle API hosting sensitive data
#*
#* @apiDescription This API contains sensitive data, the data-owner does not
#* want to share this data but does want a consultant to build a health
#* economic model using it, and wants that consultant to be able to run
#* the model for various inputs
#* (while holding certain inputs fixed and leaving them unknown).

#* Run the DARTH model
#* @serializer csv
#* @param path_to_psa_inputs is the path of the csv file containing the PSA parameters
#* @param model_functions gives the GitHub repository to source the model code
#* @param param_updates gives the replacement values of the editable parameters
#* @post /runDARTHmodel
function(path_to_psa_inputs = "parameter_distributions.csv",
         model_functions = paste0("https://raw.githubusercontent.com/",
                                  "BresMed/plumberHE/main/R/darth_funcs.R"),
         param_updates = data.frame(
           parameter = c("p_HS1", "p_S1H"),
           distribution = c("beta", "beta"),
           v1 = c(25, 50),
           v2 = c(150, 70)
         )) {


  # source the model functions from the shared GitHub repo...
  source(model_functions)

  # read in the csv containing parameter inputs
  psa_inputs <- as.data.frame(readr::read_csv(path_to_psa_inputs))

  # for each row of the data-frame containing the variables to be changed...
  for(n in 1:nrow(param_updates)){

     # update parameters from API input
     psa_inputs <- overwrite_parameter_value(
                                existing_df = psa_inputs,
                                parameter = param_updates[n,"parameter"],
                                distribution = param_updates[n,"distribution"],
                                v1 = param_updates[n,"v1"],
                                v2 = param_updates[n,"v2"])
  }

```

```r
  # run the model using the single run-model function.
  results <- run_model(psa_inputs)

  # check that the model results being returned are the correct dimensions
  # here we expect a single dataframe with 6 columns and 1000 rows
  assertthat::assert_that(
    all(dim(x = results) == c(1000, 6)),
    class(results) == "data.frame",
    msg = "Dimensions or type of data are incorrect,
  please check the model code is correct or contact an administrator.
  This has been logged"
  )

  # check that no data matching the sensitive csv data is included in the output
  # searches through the results data-frame for any of the parameter names,
  # if any exist they will flag a TRUE, therefore we assert that all = F
  assertthat::assert_that(all(psa_inputs[, 1] %in%
        as.character(unlist(x = results,
                            recursive = T)) == F))

  return(results)

}
```

Deploying an API

There are numerous providers of cloud computing services. The most convenient, yet not the cheapest, service is offered by RStudio Connect. An account is required for this, but provides the benefit of being able to deploy the API directly from the Rstudio integrated development environment. RStudio have a blog on how to publish an API created using plumber to RStudio connect here.

Interacting with the API

We first show how to run the model from an R script, calling the API and retrieving the results of the model run. We then show how to use GitHub actions to automate the process, running the R script when triggered by an event (e.g. a data-update) or a scheduled time (e.g. the 1st of each month).

Interact with the API from an RScript. We use the *POST* function from the *httr* package to query the API[12] - as shown in the code chunk below. This function requires an internet connection. We provide values for several arguments:

- *url* - the URL of the RStudio Connect server hosting the API we have created using plumber.
- *path* - the path to the API within the server URL.
- *query* & *body* - objects passed to the API in list format, with names matching the plumber function arguments.
- *config* - allows the user to specify the KEY needed to access the API.

The *content* function attempts to determine the correct format for the output from the API based upon the content type. This function ensures that the result object is a dataframe.

The script then then goes on to save the data and generate a PDF report from the outputs using the RMarkdown package[3], the code for which can be found here. The R-Markdown report uses functions adapted from the *darkpeak* R package.

Code chunk 2 - Query the API, retrieve model results and generate report **(this code is run by the consultant)**

```
# remove all existing data from the environment.
rm(list = ls())

library(ggplot2)
library(jsonlite)
library(httr)

# run the model using the connect server API
results <- httr::content(
  httr::POST(
    # the Server URL can also be kept confidential, but will leave here for now
    url = "https://connect.bresmed.com",
    # path for the API within the server URL
    path = "rhta2022/runDARTHmodel",
    # code is passed to the data-owner API from GitHub.
    query = list(model_functions =
                   paste0("https://raw.githubusercontent.com/",
                          "BresMed/plumberHE/main/R/darth_funcs.R")),
    # set of parameters to be changed ...
    # we are allowed to change these but not some others
    body = list(
      param_updates = jsonlite::toJSON(
        data.frame(parameter = c("p_HS1","p_S1H"),
                   distribution = c("beta","beta"),
                   v1 = c(25, 50),
                   v2 = c(150, 100))
      )
    ),
    # we include a key here to access the API here the key is a env variable
    config = httr::add_headers(Authorization = paste0("Key ",
                                                      Sys.getenv("CONNECT_KEY")))
  )
)

# write the results as a csv to the outputs folder...
write.csv(x = results,
          file = "outputs/darth_model_results.csv")

source("report/makeCEAC.R")
source("report/makeCEPlane.R")

# render the markdown document from the report folder,
# passing the results dataframe to the report.
rmarkdown::render(input = "report/darthreport.Rmd",
                  params = list("df_results" = results),
                  output_dir = "outputs")
```

Living HTA - scheduling model report updates. Once the API is created and hosted online, it can be called any time. The advantage of this is that any updates to either the model code, or the data used by the model, can be undertaken separately and the model re-run by either party. Calls to the API can also be scheduled at routine intervals. This would enable the health economic evaluation model report to be updated, without human interaction, at regular intervals to reflect the most up-to-date data.

In the example below we show how a GitHub Actions (other providers available) workflow can be used to automate an update to a health economic evaluation[13]. The workflow runs at 0:01 on the first day of every month or any time there are changes made to the source code. It first clones the GitHub repository on a GitHub actions Windows 2019 server, then install the necessary dependencies, before running the script described above to generate the model report. It creates a pull request to the repo with this new updated report. If GitHub is not the preferred location of report storage, it is possible to send the report via email or save to cloud storage solutions such as Google Drive or Dropbox.

Code chunk 3 - Automated report updates

```
on:
  push:
    branches:
    - main
  schedule:
    - cron: '1 1 1 * *'

name: Run DARTH model via API
jobs:
  createPullRequest:
    runs-on: windows-2019
    env:
      GITHUB_PAT: ${{ secrets.GITHUB_TOKEN }}
  # Load repo and install R
    steps:
    - uses: actions/checkout@master
    - uses: r-lib/actions/setup-r@master

    - name: Setup pandoc
      uses: r-lib/actions/setup-pandoc@v2
      with:
        pandoc-version: '2.17.1.1'

    - name: Install TinyTeX
      uses: r-lib/actions/setup-tinytex@v2
      env:
          # install full prebuilt version
          TINYTEX_INSTALLER: TinyTeX

    - name: Install dependencies
      run: |
          install.packages(
          c("reshape2", "jsonlite", "httr", "readr", "rmarkdown", "markdown")
          )
          install.packages(
          "scales", dependencies = TRUE, repos = 'http://cran.rstudio.com/'
          )
          install.packages(
          "ggplot2", dependencies = TRUE, repos = 'http://cran.rstudio.com/'
          )
      shell: Rscript {0}

    - name: Run the model from API and create report
      env:
         CONNECT_KEY: ${{secrets.PLUMBER_SECRET}}
      run: |
          source("scripts/run_darthAPI.R")
      shell: Rscript {0}

    - name: Create Pull Request
      uses: peter-evans/create-pull-request@v3
      with:
        token: ${{ secrets.GITHUB_TOKEN }}
        commit-message: Automated Model Run from API
        title: 'Living HTA Automated Model Run'
        body: >
          Automated model run
        labels: report, automated pr
```

Results

All source code for the API, the economic model, the automated model update framework, and the example dataset are available online (see *Software availability*[6] and *Underlying data*[14]).

The most up to date automated report, based on the data held on the exemplar API (hosted on RStudio Connect), can always be found here.

The method has been validated by two co-authors using Windows and MAC with example data (see *Underlying data*[14]). Those validating the method were able to run the model with updated parameter values without access to sensitive data, were able to trigger the automated report generation based on existing sensitive data, and were able to query the model through an example R-Shiny application, hosted on GitHub (see *Software availability*[6]). However we are keen to validate the method further, and invite collaboration. A live exemplar API is currently hosted by Lumanity (using the exact source code provided open access). If the reader is interested to test the functionality of the API please contact the corresponding author, who can provide the key.

Discussion

As the collection and storage of large data sets has become more commonplace in health & health care settings, this data is increasingly being used to inform decision making. However, concerns about the security of this data, and the ethical implications about linked data sets, make the owners of this valuable resource particularly reluctant to share data with health economic modelling teams. The ability to host APIs on data-owners' servers, and send the model to the data rather than the data to the model, is one potential solution to this problem. The example described in this paper may be relatively simple, but gives a tech savvy health economist everything they need to set up a modelling framework which does not rely on the sharing of data by a pharmaceutical company (or other *data-owner*).

The framework described has a number of benefits.

- Firstly, no data needs to leave the data-owner's server. This is likely to significantly reduce administrative burden for both the *data-owner* and the *consultant*, and reduce the number of data-leaks.

- Separating the data from the model has significantly improved the transparency of the health economic model. Allowing others to critique methods & hidden structural assumptions, test the code and identify bugs should improve the quality of models in the long run. It also enables the pool of people working on developing the health economic model and accompanying user-interface to be widened, without concern for confidentiality & data security. For example a shiny application could be developed for a model built under this framework without the programmer needing access to any sensitive data or information. However, if it is necessary to restrict access to model code, it is possible for the API to be passed 'private' source code. As keen proponents of open source modeling that was determined to be beyond the scope of this publication.

- The computational burden of the model is handled on a remote server. The power of these servers is typically considerably greater than that of a typical personal computer, speeding up model run time considerably. This is likely to be especially important for models that incorporate uncertainty through monte-carlo sampling algorithms which can be parallelized on machines with multiple cores[15], for example probabilistic one way sensitivity analysis[16] or partial expected value of perfect information[17].

- The use of APIs to perform distinct tasks can improve interoperability within the field of health economics. Different modules, or tasks within a modelling framework can be written in different languages (e.g. R, Python, Julia & C++) and linked using APIs. This is likely to improve collaboration between different sub-disciplines, which often use different languages (e.g. health economists in R and data-scientists in Python).

- API calls can be made at any time, and will always reflect the data held by the data owner. In many cases these datasets are updated regularly, allowing companies, and other stakeholders, to see the results of the decision model based on the most up to date data, without needing human intervention to: send new datasets, re-run analysis, write a report, and provide that report in a suitable format for the data-owner. Automating model updates at set schedules, or when data is updated, may be invaluable where data is updated regularly, as has been the case throughout the COVID-19 pandemic.

- Any model can be passed to the API, as long as the inputs and outputs to the model meet the requirements of the API. This means that multiple health economic models could be passed to the API, to be run using the data on the data-owner's server, and compared to account for structural uncertainty.

- In the (intentionally simple) example we give in this paper, it is assumed that the data owner provides readily estimated parameter values. In a real world use case, this does not need to be the case: the consultant may also send a survival model specification, for example, to estimate parameters from individual patient data, before the parameters are then passed to the health economic model.

- Specifying a model without access to and feedback from the actual data may come with its own challenges. However, if these hurdles are overcome, the benefit is that it enforces a thorough statistical analysis plan by default, in every use case. This helps to avoid biases introduced by stakeholder incentives.

However, the framework has a number of limitations:

- Firstly, the method is relatively complex, and requires a strong understanding of health economic modelling in R, API creation and hosting, RMarkdown or other automated reporting packages, and GitHub Actions. While we hope that this paper provides a useful resource to health economists seeking to utilise these methods, the bulk of the industry still operates in MS Excel[18]. Providing tuition to upskill health economists, or creating teams consisting of both health economists and data-scientists and software engineers may mediate this limitation somewhat. Groups like the R for HTA consortium has the potential to play a crucial role in upskilling the industry.

- There are still likely to be concerns about data security, even with the authentication procedures built in to the API functionality. Collaboration with experts in this field may mediate this significantly, since there is no fundamental reason why health data is any more sensitive, or vulnerable, than the plethora of other data (including banking data) that relies on APIs every day. It will be important to reassure companies that the use of APIs is likely to reduce, not increase the risk of data breaches, and that every interaction with the data can be logged.

- There is a risk that running the model remotely will result in the perception that the model is a 'black box'. The use of user-interfaces (such as those increasingly being created in *shiny*) to interrogate the model, as well as the increased transparency associated with being able to share code on sites such as GitHub, should reassure stakeholders that this framework is more transparent than the existing spreadsheet based solutions[19].

- R is single threaded, and therefore will only work on one task at a time. This can make it slow when lots of requests are made simultaneously, which may occur if a model takes a long time to run. Most hosting platforms (including RSConnect, where we have deployed this example) solve this problem by creating multiple R processes, which work in isolation. The PRISM solution outlined by Abidi *et al.*[2] uses OpenCPU which works in the same way. This is fine for the example model we provide, since our model doesn't store any information required by other users, who would be working on another R process, during the session. Further information can be found on the plumber guide here: https://www.rplumber.io/articles/hosting.html.

- Often, when building a model, it is helpful to have the underlying data to be able to investigate the data, often through the generation of descriptive statistics. The process of sharing pseudo-data enables modellers to ensure that the models they create conform to the structure of the data input. However, the modeller still needs to be able to write code that is versatile enough to cope with data with unknown distributions ranges and number of observations. This is easily solved, again by improved training and the use of standard packages such as *hesim* and *heemod*[20,21].

The recent working paper by Adibi *et al.*[2] has provided a similar call to action, extolling the virtues of the API for decision modelling, and showing how APIs can be used to shift much of the computational burden away from those querying models, making models more accessible. However, there are several limitations to this innovative paper. Firstly, while the authors outline a framework for making models more transparent and accessible, and describe how they have done this for a number of models using the PRISM server, they do not provide instruction on how to replicate this process. Additionally, while the authors state that "A practical model accessibility platform should be able to protect confidential information such as patient data and confidential pricing" (p6), the framework as described would require companies to give the owners of the model accessibility platform access to their confidential data, or else host the model accessibility platform themselves.

This paper has attempted to address some of these limitations, providing open source code for the creation and deployment of an API with an accompanying automated health economic evaluation update framework. It also provides open source code on two new pieces of additional functionality not previously described elsewhere; firstly it demonstrates how companies can host APIs themselves to negate the need to share data with subject experts, and secondly it demonstrates how model updates can be automated with scheduled workflows run on remote servers.

Conclusions

This example framework, with accompanying open source code base, demonstrates that it is possible, within a HEOR setting, to separate a health economic model from the data, and automate the main steps of the analysis pipeline. We believe this is the first application of this procedure in the HEOR context, and is certainly the first example to be made open source for the benefit of the wider community. We hope that this framework will improve the transparency of health economic models, reduce the cost and administrative burden of updating models, and increase the speed at which updates can occur.

Data availability

Underlying data

Zenodo: Parameter distributions. https://doi.org/10.5281/zenodo.6727629[14].

This project contains the following underlying data:

- parameter_distributions.csv (example dataset for modification. This dataset also sits within the server, with some of the rows marked as non-editable; these are characterised as 'sensitive' throughout the manuscript. This dataset is edited in 'Code Chunk 2' to test the API).

Software availability

Source code available from: https://github.com/RobertASmithBresMed/plumberHE.

Archived source code at time of publication: https://doi.org/10.5281/zenodo.6556888[6]

License: MIT

Acknowledgements

We would like to thank the participants at R-HTA Oxford, Richard Birnie (Lumanity) and Dawn Lee (Lumanity) for feedback on the original concept and manuscript. All errors are the fault of the authors.

References

1. Williamson EJ, Walker AJ, Bhaskaran K, *et al.*: **Factors associated with covid-19-related death using opensafely.** *Nature.* 2020; **584**(7821): 430–436.
PubMed Abstract | Publisher Full Text | Free Full Text
2. Adibi A, Harvard S, Sadatsafavi M: **Programmable interface for statistical & simulation models (prism): Towards greater accessibility of clinical and healthcare decision models.** *arXiv preprint arXiv: 2202.08358.* 2022.
Reference Source
3. Xie Y, Dervieux C, Riederer E: **R Markdown Cookbook.** Chapman and Hall/CRC, Boca Raton, Florida, 2020.
Reference Source
4. Chang W, Cheng J, Allaire JJ, *et al.*: **shiny: Web Application Framework for R.** 2021.
Reference Source
5. Smith R, Schneider P: **Making health economic models Shiny: A tutorial [version 2; peer review: 2 approved].** *Wellcome Open Res.* 2020; **5**: 69.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Smith R, Schneider P, Mohammed W: **Robertasmithbresmed/plumberhe: R-hta june 2022.** software, 2022.
http://www.doi.org/10.5281/zenodo.6556888
7. Alarid-Escudero F, Krijkamp EM, Enns EA, *et al.*: **A tutorial on time-dependent cohort state-transition models in r using a cost-effectiveness analysis example.** *arXiv preprint arXiv: 2108.13552.* 2021.
Publisher Full Text
8. Alarid-Escudero F, Krijkamp E: **DARTHgit/cohort-modeling-tutorial-intro: Second Zenodo release.** 2022.
Publisher Full Text
9. Ooms J: **The jsonlite package: A practical and consistent mapping between json data and r objects.** *arXiv:1403.2805 [stat.CO].* 2014.
Reference Source
10. Schloerke B, Allen J: **plumber: An API Generator for R package**

version 1.1.0, 2021.
Reference Source

11. Wickham H, Danenberg P, Csárdi G, *et al.*: **roxygen2: In-line documentation for r.** *R package version.* 2020; **7**(1).
12. Wickham H: **httr: Tools for Working with URLs and HTTP.** R package version 1.4.2. 2020.
Reference Source
13. Chandrasekara C, Herath P: **Introduction to github actions.** In: *Hands-on GitHub Actions.* Springer, 2021; 1–8.
Publisher Full Text
14. Smith R: **Parameter distributions.** dataset. 2022.
http://www.doi.org/10.5281/zenodo.6727629
15. R Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2020.
Reference Source
16. McCabe C, Paulden M, Awotwe I, *et al.*: **One-Way Sensitivity Analysis for Probabilistic Cost-Effectiveness Analysis: Conditional Expected Incremental Net Benefit.** *Pharmacoeconomics.* 2020; **38**(2): 135–141.
PubMed Abstract | Publisher Full Text | Free Full Text
17. Brennan A, Kharroubi S, O'hagan A, *et al.*: **Calculating partial expected value of perfect information via monte carlo sampling algorithms.** *Med Decis Making.* 2007; **27**(4): 448–470.
PubMed Abstract | Publisher Full Text
18. Incerti D, Thom H, Baio G, *et al.*: **R you still using excel? the advantages of modern software tools for health technology assessment.** *Value Health.* 2019; **22**(5): 575–579.
PubMed Abstract | Publisher Full Text
19. Pouwels XGLV, Sampson CJ, Arnold RJG, *et al.*: **Opportunities and Barriers to the Development and Use of Open Source Health Economic Models: A Survey.** *Value Health.* 2022; **25**(4): 473–479.
PubMed Abstract | Publisher Full Text
20. Incerti D, Jansen JP: **hesim: Health economic simulation modeling and decision analysis.** *arXiv preprint arXiv:2102.09437.* 2021.
Publisher Full Text
21. Filipovic-Pierucci A, Zarca K, Durand-Zaleski I: **Markov models for health economic evaluation modelling in r with the heemod package.** *Value Health.* 2016; **19**(7): A369.
Publisher Full Text

Part 5 - Discussion and conclusion

Chapter 9. General discussion and conclusion

Main findings

The focus of this thesis is on exploring the balance between the external-validity and usability of models used in health economic evaluations, with a specific emphasis on models related to physical activity. The thesis has identified ways in which existing models can be adapted to make them more externally valid (Ch. 3 & 4), a new model developed to address existing methodological limitations which impact external validity (Ch. 6), and methods developed to help improve the usability of models (Ch 7 & 8). The aim throughout has been to encourage the development of models that are the optimal combination of externally valid and usable to maximise the public health benefits of health economic evaluations.

The first chapter outlined the benefits of physical activity, the strategy of public health institutions, challenges in surveillance and measurement, and the determinants and correlates of physical activity.

Chapter 2 reviewed the methods used for estimating the cost-effectiveness of physical activity interventions, with an emphasis on the techniques employed by models submitted to NICE. The review was originally carried out in 2017 but has been updated in 2023. It identified several limitations of these methods, such as the use of categorical representations of physical activity (e.g. active/inactive), the use of discrete or linear dose-response relationships, a lack of inter-sectoral analysis, limited attention to the duration and decay of intervention effects, and a narrow focus on a subset of health conditions related to physical activity. Many of these limitations were linked to the choice of a simple model structure (Markov models) commonly used in HTA submissions for pharmaceutical products but which make it very difficult to overcome these limitations. The thesis also notes that submitted models tend to have limited transparency beyond the technical report and were difficult to access, both for laypeople and researchers without direct communication with the authors. Finally, although there was a shift toward the use of QALYS, many studies in the wider literature did not use a cost-utility framework and instead looked at cost-consequence.

Chapters 3, 4 and 5 attempted to quantify the effect of some of the structural assumptions used in existing models and suggest methodological adaptations to better match the understanding of the epidemiological literature in Chapter 1, and to address some of the modelling limitations identified in Chapter 2.

Chapter 3 evaluated the impact on net monetary benefit of the use of the VSLY approach (vs the existing VSL approach) for valuing deaths averted in HEAT and found that while it could improve the external validity of estimates for interventions targeting non-representative groups, such as those in or near educational institutions, it may also decrease the tool's ease of use and acceptance among key stakeholders. Importantly, it raised ethical questions around the use of the tool to make resource allocation decisions between countries with different mortality rates, VSL estimates and demographics.

Chapter 4 assessed the implications of incorporating a non-linear dose-response function into the HEAT model and found that it led to significantly different estimates for populations with non-representative levels of baseline physical activity, which may be important when targeting specific subgroups, such as cycle lane users. Understanding the implications of the use of a different dose-response function within and between countries is important for both overall outcomes but also inequalities within and between countries.

Chapter 6 tested the feasibility of a method to incorporate a consideration of physical activity tracking over the life course, incorporating multiple conditions simultaneously with age specific risks and utilising data on baseline population quality of life to inform a utility model. It also tested the effect of several common structural assumptions made in health economic models of physical activity including assumptions around intervention effect duration and methods of utility estimation in the presence of comorbidity.

Chapters 7 and 8 explore ways to enhance the transparency, usability, and efficiency of health economic models in the form of tutorial papers with accompanying open-source code and software prototypes.

Chapter 7 outlined a method by which a user interface can be built and deployed online for an economic model built in a script-based programming language to improve access to, transparency of and engagement with models. It is posited that web-based user interfaces can help to improve the usability of health economic models without sacrificing external-validity, making an existing model more accessible and useful for a wider range of stakeholders.

Chapter 8 extends the methods discussed in Chapter 7 outline how health economic evaluations can be adapted to be ‘Living’ so as to avoid the information and data on which they are based becoming out-of-date, a common reason why models become less usable. The method applies new methods from data-science to health economics, to update the model report as new effectiveness evidence is published, new prices or discounts are set, or where the model is directly informed by analysis which itself can be continually updated. This has huge potential in the context of public health and physical activity more specifically, since data is regularly collected on patient outcomes by General Practitioners and could be accessed via NHS Digital, allowing the cost-effectiveness of interventions to be continually reviewed over time.

Throughout the thesis, the focus has remained on achieving the right balance between usability and external validity. This can be aided by iteration in model building, incrementally adapting and improving models over time. This has the added benefit of potentially bringing decision-makers along on a journey over time, as has been the case with the HEAT, to gradually improve model external-validity while maintaining usability.

Contributions to existing knowledge

Consideration of potential trade-offs between external-validity and usability.

Little consideration has been given to the trade-off between the external-validity and usability of health economic models in the literature. This work hopes to contribute to the literature on this topic, providing a discussion throughout about potential pitfalls that some models may face which limit their ability to have impact on decisions. I provide some examples in which there does appear to exist a clear trade-off between external-validity and usability and outline some ways in which this could be mitigated, and some ways in which one can be incrementally improved with limited negative repercussions on the other.

Valuation methods for reductions in mortality rates due to physical activity

Chapter 3 investigates the impact of the choice of valuation for mortality reduction due to physical activity, using the HEAT for 44 countries as a case study. The choice of the valuation method, valuing life years saved using the VSLY or valuing deaths averted using the VSL, can have a profound impact on the net monetary benefit estimates derived from physical activity models. This is especially the case where populations impacted by an intervention are demographically unrepresentative of the general population in a country. As a result of the limitations of these approaches a decision was made to use QALYs in the analysis for chapter 6.

Dose-response functions for physical activity

The scoping review in Chapter 2 identified that a common limitation of physical activity models was the simplification of the dose-response relationship between physical activity and health outcomes (including mortality) to be either categorical (with physical activity split into categories such as Active/Inactive), or linear with a maximum effect. Chapter 4 investigates the impact of the choice of dose-response function for the relationship between physical activity and mortality using the HEAT as a case study. It showed that the net monetary benefit associated with increases in physical activity were highly sensitive to the choice of dose response relationship, and that this sensitivity was starker for interventions which affect groups which are particularly inactive or active compared to the general population. Chapter 6 outlined a method of incorporating independent non-linear dose response functions for multiple diseases simultaneously, estimating incidence and prevalence given a set of

assumed disease durations. It incorporated uncertainty around the dose response relationship, incidence and disease duration into the model by sampling from distributions around each for a large number of iterations (probabilistic sensitivity analysis).

Feasibility of incorporating long-run tracking of physical activity over the life course, and impact of assumptions around intervention effect decay

Chapter 6 also incorporated long-run tracking of physical activity over the life course via estimations of the distribution of physical activity at each age. While the simplifying assumption that individuals remain at the same percentile is unrealistic and should be varied if future studies wish to investigate inequalities, the result is that the overall population estimates are credible. This is important because changes in habit formation that last a lifetime should be understood in the context of falling general levels of PA. The HEAT model, discussed in Chapters 3-5, does not incorporate changes in physical activity levels over time. Finally, assumptions around intervention effect decay were varied substantially in Chapter 6. This analysis highlights how sensitive models of physical activity intervention are to assumptions around duration of effect, and extent of decay. While the model discussed in Chapter 6 is particularly sensitive due to the intervention group (children) and the nature of the diseases included (NCDs), many other models have previously been developed with the similar aged intervention groups and diseases and have made varying assumptions on the decay in intervention effect.

Tutorial on the development of user-interfaces for health economic models

One of the limitations of models identified in Chapter 2 was the limited ability to engage with the models, other than reading the reports. Chapter 7 outlines a method by which user-interfaces can be built for health economic models constructed in R, deployed online and shared with stakeholders to enable them to better engage with and understand the model. This chapter is written in tutorial format, guiding the reader through the process of building the user-interface and deploying it, with the aim of helping health economists to make their models more transparent and usable.

A prototype 'Living' health economic evaluation model report

Recent advances in data-science and computing make it increasingly feasible for some parts of health economic model updates to be automated. Chapter 8 provides a description of the process by which a health economic model could be automatically updated as data, stored on a secure server, is updated, or when triggered by a user of a UI (as developed using Chapter 7). There is a lot of research ongoing on the wider context for Living Health Technology Assessment, of which interventions to increase physical activity may be one but having a published methodology for Living Health Economic Evaluation with prototype application and open-source code goes some way to demonstrating that the process is at least technically possible.

Implications

Trade-offs

Understanding the trade-offs between the external-validity and usability of health economic models should be prioritised when conceptualising a model. Increasing the external validity of a model at the cost of usability may not be desirable - despite being of academic interest.

Iteration is useful. Incrementally improving the external-validity, and usability, of models over time allows decision makers and the broader stakeholder community to engage with the process over time and gain confidence. For this to work best, model code should be made open source, to enable others to test the implications of adaptations to models, and apply to the model 'owners' with suggestions, with the aim of gradually improving the model over time - an approach that is preferential to a preponderance of models.

External validity

Economic models which estimate the benefits of physical activity should carefully consider the implications of using categories of physical activity or assuming a linear dose-response relationship. The benefit of physical activity has been shown to be both continuous and non-linear, characterised by the WHO's guidance that 'some physical activity is better than doing none' and even at higher levels of activity 'More physical activity is better' (Bull et al., 2020). The work discussed in Chapter 4 also showed that model results are highly sensitive to this choice, especially when populations affected by an intervention are non-representative of the general population.

Models which incorporate societal valuations on changes in mortality rates should consider whether they wish to value deaths-averted, or life-years gained. Chapter 2 showed that the valuation of deaths-averted is simpler to calculate but can result in particularly high valuations on interventions which reduce the risk of death in older persons - which is of particular concern when the intervention being assessed is targeted at populations which are demographically unrepresentative of the general population.

A similar theme emerges from Chapter 6. In the absence of data with life-long follow up, those building health economic models of physical activity interventions should consider a feasible duration over which to decay the effect of the intervention. Decision makers should be sceptical of models without decay in effect. Modellers should also consider how to incorporate long-term benefits into models, and decision-making bodies should consider the implications of their choice of discount rates for public health interventions, especially where discounting the future too heavily may result in a significant healthcare burden for later generations. Finally, sensitivity analysis should be undertaken throughout to ensure that decision makers are easily able to determine which factors the decision is particularly sensitive to.

Usability

Models built in script-based programming languages, such as R, can be rapidly deployed online and shared with stakeholders worldwide via a user-interface. Although there are initial costs in the development of a user-interface, the long-term benefits of utilising the methods outlined in Chapter 7 are likely to far exceed costs in most cases. Since the publication of the tutorial paper, the paper has been cited in several other publications calling for the utilisation of the method in health economic models more generally (Naylor et al. 2023; Xin, 2022; Pouwels et al. 2022). However, to get the greatest benefit it is best to engage with decision makers and other stakeholders early to ensure that the UI meets their requirements in both functionality and aesthetics, since usability is determined by the user and whether the model is used to inform policy is often dependent on engagement with the decision-maker.

The technical capability exists to facilitate 'Living Health Economic Evaluation' to ensure that health economic evaluation reports always reflect the latest available data (Chapter 8). The implications of this are less certain, since the technical feasibility is just one consideration that must be considered when determining whether 'Living Health Economic Evaluation' is desirable (Thokala et al. 2023). However, it is likely most useful when the model depends on data and evidence that is continually being updated, and when decisions need to be made rapidly. It is therefore most obviously useful for commissioning in infectious diseases. However, the process may also be useful in informing commissioning decisions where the intervention effect is highly uncertain, but evidence is continually being published. One example would be a physical activity intervention which has a limited evidence base but for which multiple trials are underway.

Strengths & Limitations

The individual strengths and limitations of each study are discussed independently within Chapters 3 - 8. However, there are several overarching strengths and limitations of the thesis.

Strengths

Taking a two-year gap during my PhD to work on the UK government's response to the pandemic has had its drawbacks, but it has also provided me with significant benefits. One major advantage is that I have been able to publish the work in peer-reviewed journals and present methods at (online) conferences and through short course tutorials and have made all of the source code related to these studies open-source. As a result, much of the work included in this thesis has been influenced by or benefited from feedback from colleagues from a variety of backgrounds, including academia, government, industry, and consulting. Additionally, this gap has allowed some of the work to be replicated (ReproHack Hub Team; 2020; 2021.) and built upon by other academics (Smith et al. 2022b; Haake et al. 2022), and has had a direct impact on policy and methods used in industry (see

Contributions section) in ways that may not have been possible within the typical three-year PhD timeframe.

The research both adapts the HEAT model in two ways and develops a new health economic model from scratch. Testing changes to an existing model is useful in the context of incrementally editing modelling methods while holding all other structural assumptions constant to determine the effect of the assumption in isolation, thereby assessing the marginal impact of the adaptation on external-validity and usability, with a discussion about the relative merits of the adaptation. Building a new model allows for the use of a microsimulation model which avoids many of the limitations of Markov models identified in the review chapters. It is a more ambitious approach, but one that allows for analysis on the effect of several distinct methodological approaches on the cost-effectiveness of interventions targeted at children and adolescents. As a result, the thesis provides a useful reference point to someone attempting to build or adapt health economic models for interventions targeted at increasing population physical activity.

The latter part of the thesis presents tutorials on two new methods that combine new techniques from data science and computing to health economic evaluation. These are to date the only published papers that provide information on the benefits of, outline the concept, and provide detailed instructions including all source code to enable others to apply the methods to improve the usability of their own models. This is a strength because it enables the work to have considerable impact beyond this thesis, via the improvement in the usefulness of health economic models developed using the methods. The papers have been written for a general audience to facilitate their use in decision models for public health and health technology assessment for pharmaceuticals.

Limitations

This thesis has several limitations that should be acknowledged. Firstly, a comprehensive systematic review of the literature was not conducted. Instead, the focus in Chapter 2 was on NICE models and papers reviews identified by a review of reviews. Future research could examine the limitations identified in a larger number of model reports directly. However, it was deemed beyond the scope of this thesis to conduct a full systematic review, and the limitations identified by other reviews were considered sufficient to justify the choice of structural assumptions to undertake sensitivity analysis on. Additionally, while the PACEM model in Chapter 6 addresses some limitations of previous physical activity models identified in Chapter 2, it does not address the intersectoral consequences of physical activity policy or include any consideration of distributional effects, although the model is set up to allow for differential impact by deprivation quintile if this information is available in the future. These limitations are primarily the result of narrowing the scope of the thesis and the impact of other responsibilities during the pandemic on time and resource constraints.

The PACEM model, described in Chapter 6, was not developed in consultation with decision-makers. While the original conceptual model was developed with a group of experts including individuals working

within government departments, they did not have subsequent input due to other commitments during the pandemic. As a result, it does not provide direct policy recommendations and instead makes only methodological contributions to the literature. It is therefore of primary interest to those attempting to better understand the implications of structural assumptions in cost-effectiveness models, rather than decision-makers. However, all code is made open-source so the model could be adapted to a specific decision problem in the future. This is in contrast to the body of work developed in collaboration with parkrun, for which the decision problem was formulated by the decision maker, and the papers published on the WHO's HEAT for walking and cycling which has been widely used and tailored to the needs of decision-makers over a decade. It is also a weakness relative to the published literature discussed in Chapter 2, much of which was developed as part of NICE appraisals and therefore has the benefit of direct input from decision-makers and other stakeholders, as well as a clear route to impact via NICE guidance.

The tutorial on the development of web-based user-interfaces for health economic models has been well received. However, it is uncertain whether the methods discussed in the Living Health Economic Evaluation section can be implemented in ongoing projects. The field of Living HTA is still relatively new, and further research needs to be done to better understand the needs of decision-makers, technology developers and the general public in their engagement with health economic evaluation before it can be used in standard practice.

Areas for future research

There is a growing body of research on how physical activity levels change over the life course. To better understand these trajectories and their determinants, it would be helpful to improve the availability of high-quality longitudinal studies over the entire life course with consistent, and ideally objective, measurements of physical activity. Additionally, a comprehensive review of the methods used to include long-term physical activity data in health economic models would be helpful.

The PACEM model could be improved by including subgroup analysis and by using data on physical activity and disease rates by levels of socioeconomic deprivation. This would enable the evaluation of the distributional cost-effectiveness (Asaria et al. 2016) of physical activity interventions, which is particularly important given that physical activity levels and engagement with public health interventions can vary significantly between socioeconomic groups (Smith et al. 2020b). Furthermore, the model could be adapted to different regions in the UK to account for variations in physical activity, socioeconomic deprivation, and health outcomes. Also, the model could be modified to consider feedback loops and interdependencies between physical activity, Type 2 Diabetes, Ischemic Stroke, and other health conditions.

Additional research could explore ways to improve the ease of use and user-experience of a web-based health economic model for different stakeholders. In particular, research could examine how decision-

makers prioritise external-validity and ease of use in the HEAT model, with the goal of developing a tool that offers different levels of complexity for different users. In the future, A-B testing (a method of comparing two versions of a product or website to determine which performs better) and direct feedback within the tool could be used to collect quantitative data on user preferences.

The concept of "*Living HTA*" and "*Living Health Economic Evaluation*" is emerging in health economics (Thokala et al. 2023). As the allocation of government funds becomes increasingly constrained, the demand for healthcare continues to rise, and new medical technologies and treatments become available, it is necessary to explore new ways to fund healthcare. One promising area for future research is the integration of data from healthcare services and individual patient records into health economic models, which would enable policymakers to evaluate the cost-effectiveness of interventions in real-time and make informed decisions about commissioning.

Conclusion

The thesis aimed to examine the trade-off between the external-validity and usability of models used for health economic evaluations, with a specific focus on physical activity models. The research identified ways to adapt existing models to increase external validity, develop new models to address methodological limitations, and improve the usability of models. By doing so, this body of research aimed to encourage the development of models that are the optimal combination of externally valid and usable, to improve their value to decision makers and thereby maximise the public health benefits of health economic evaluations.

References

Abu-Omar, K., Rütten, A., Burlacu, I., Schätzlein, V., Messing, S. and Suhrcke, M., 2017. The cost-effectiveness of physical activity interventions: a systematic review of reviews. Preventive medicine reports, 8, pp.72-78.

Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplaincourt, P.O. and Jacobs, D.R., 2000. Compendium of physical activities: an update of activity codes and MET intensities. Medicine and science in sports and exercise, 32(9; SUPP/1), pp.S498-S504.

Aittasalo, M., Miilunpalo, S. and Suni, J., 2004. The effectiveness of physical activity counseling in a work-site setting: A randomized, controlled trial. Patient Education and Counseling, 55(2), pp.193-202.

Alarid-Escudero, F., Krijkamp, E., Enns, E.A., Yang, A., Hunink, M.M., Pechlivanoglou, P. and Jalal, H., 2023. An introductory tutorial on cohort state-transition models in R using a cost-effectiveness analysis example. Medical Decision Making, 43(1), pp.3-20.

Alarid-Escudero, F., Krijkamp, E.M., Pechlivanoglou, P., Jalal, H., Kao, S.Y.Z., Yang, A. and Enns, E.A., 2019. A need for change! A coding framework for improving transparency in decision modeling. Pharmacoeconomics, 37(11), pp.1329-1339.

Allender, S., Hutchinson, L. and Foster, C., 2008. Life-change events and participation in physical activity: a systematic review. Health promotion international, 23(2), pp.160-172.

Anderson L, Schnohr P, Schroll M, Hein H. 2000. All-cause mortality associated with physical activity during leisure time, work, sports, and cycling to work. Arch. Intern. Med; 160:1621-1628.

Anokye N, Lord J, Fox-Rushby J, 2013. Physical activity: brief advice for adults in primary care. London: Nice public health guidance, 44. Available from https://www.nice.org.uk/guidance/ph44/evidence/economic-modelling-of-brief-advice-on-physical-activity-for-adults-in-primary-care-pdf-430312141

Ara, R. and Brazier, J.E., 2010. Populating an economic model with health state utility values: moving toward better practice. Value in Health, 13(5), pp.509-518.

Ara, R. and Brazier, J.E., 2011. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in Health, 14(4), pp.539-545.

Ara, R.M. and Brazier, J. 2017. Estimating Health State Utility Values for Comorbidities. PharmacoEconomics, 35 (Supp 1). pp. 89-94. ISSN 1170-7690.

Asaria, M., Griffin, S. and Cookson, R., 2016. Distributional cost-effectiveness analysis: a tutorial. Medical Decision Making, 36(1), pp.8-19.

Baio, G. and Heath, A., 2017. When simple becomes complicated: why Excel should lose its place at the top table. Global & Regional Health Technology Assessment, 4(1), pp.grhta-5000247.

Baio, G., Berardi, A. and Heath, A., 2017. Bceaweb: A user-friendly web-app to use bcea. In Bayesian cost-effectiveness analysis with the R package BCEA (pp. 153-166). Springer, Cham.

Balzer, K., Bremer, M., Schramm, S., Lühmann, D. and Raspe, H., 2012. Falls prevention for the elderly. GMS health technology assessment, 8.

Barendregt, J.J., Van Oortmarssen, G.J., Vos, T. and Murray, C.J., 2003. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. Population health metrics, 1(1), pp.1-8.

Bates, S.E., 2021. Incorporating Psychological Mechanisms of Action in a Health Economic Model of Obesity (Doctoral dissertation, University of Sheffield).

Bates, S.E., Thomas, C., Islam, N., Ahern, A.L., Breeze, P., Griffin, S. and Brennan, A., 2022. Using health economic modelling to inform the design and development of an intervention: estimating the justifiable cost of weight loss maintenance in the UK. BMC public health, 22(1), pp.1-12.

Bauman, A., Bull, F., Chey, T., Craig, C.L., Ainsworth, B.E., Sallis, J.F., Bowles, H.R., Hagstromer, M., Sjostrom, M. and Pratt, M., 2009. The international prevalence study on physical activity: results from 20 countries. International journal of behavioral nutrition and physical activity, 6(1), pp.1-11.

Bauman, A.E., Reis, R.S., Sallis, J.F., Wells, J.C., Loos, R.J., Martin, B.W. and Lancet Physical Activity Series Working Group, 2012. Correlates of physical activity: why are some people physically active and others not?. The lancet, 380(9838), pp.258-271.

Bauman, A.E., Sallis, J.F., Dzewaltowski, D.A. and Owen, N., 2002. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. American journal of preventive medicine, 23(2), pp.5-14.

Beale, S.J., Bending, M.W., Trueman, P. and Naidoo, B., 2012. Should we invest in environmental interventions to encourage physical activity in England? An economic appraisal. The European Journal of Public Health, 22(6), pp.869-873.

Ben-Shlomo, Y. and Kuh, D., 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. International journal of epidemiology, 31(2), pp.285-293.

Bending, M., Beale, S. and Hutton, J., 2008. An economic analysis of workplace interventions that promote physical activity. PHIAC Report. York, England: York Health Economics Consortium. Available from https://www.nice.org.uk/guidance/ph13/evidence/economic-modelling-report-pdf-369939277

Blair, S.N., Cheng, Y. and Holder, J.S., 2001. Is physical activity or physical fitness more important in defining health benefits?. Medicine & Science in Sports & Exercise, 33(6), pp.S379-S399.

Breeze, P., Squires, H., Chilcott, J., Stride, C., Diggle, P.J., Brunner, E., Tabak, A. and Brennan, A., 2016. A statistical model to describe longitudinal and correlated metabolic risk factors: the Whitehall II prospective study. Journal of Public Health, 38(4), pp.679-687.

Breeze, P.R., Thomas, C., Squires, H., Brennan, A., Greaves, C., Diggle, P., Brunner, E., Tabak, A., Preston, L. and Chilcott, J., 2017. Cost-effectiveness of population-based, community, workplace and individual policies for diabetes prevention in the UK. Diabetic Medicine, 34(8), pp.1136-1144.

Breeze, P.R., Squires, H., Ennis, K., Meier, P., Hayes, K., Lomax, N., Shiell, A., Kee, F., de Vocht, F., O'Flaherty, M. and Gilbert, N., 2023. Guidance on the use of complex systems models for economic evaluations of public health interventions. *Health Economics*.

Brennan, A., Blake, L., Hill-McManus, D., Payne, N., Buckley-Woods, H., Blank, L. 2012 National Institute for Health and Clinical Excellence Public Health Walking and Cycling: local measures to promote walking and cycling as forms of travel or recreation: Health economic and modelling report. Sheffield. Available from https://www.nice.org.uk/guidance/ph41/evidence/economic-modelling-report-pdf-430262893

Briggs, A., Sculpher, M. and Claxton, K., 2006. Decision modelling for health economic evaluation. Oup Oxford.

Brennan, A., Chick, S.E. and Davies, R., 2006. A taxonomy of model structures for economic evaluation of health technologies. Health economics, 15(12), pp.1295-1310.

Brown, V., Diomedi, B.Z., Moodie, M., Veerman, J.L. and Carter, R., 2016. A systematic review of economic analyses of active transport interventions that include physical activity benefits. Transport Policy, 45, pp.190-208.

Bull, F.C., Al-Ansari, S.S., Biddle, S., Borodulin, K., Buman, M.P., Cardon, G., Carty, C., Chaput, J.P., Chastin, S., Chou, R. and Dempsey, P.C., 2020. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. British journal of sports medicine, 54(24), pp.1451-1462.

Bull, F.C., Armstrong, T.P., Dixon, T., Ham, S., Neiman, A. and Pratt, M., 2004. Physical inactivity. Comparative quantification of health risks global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization, pp.729-881.

Bull, FC. and the Expert Working Groups. Physical Activity Guidelines in the U.K.: Review and Recommendations. School of Sport, Exercise and Health Sciences, Loughborough University, May 2010.

Campbell, F., Holmes, M., Everson-Hock, E., Davis, S., Buckley-Woods, H., Anokye, N., Tappenden, P. & Kaltenthaler, E. A systematic review and economic evaluation of exercise referral schemes in primary care: A short report. Health Technol Assess 2013. Available from: https://www.nice.org.uk/guidance/ph54/evidence/evidence-review-1-a-systematic-review-and-economic-evaluation-a-short-report-pdf-4417445773

Candio, P., Meads, D., Hill, A.J. and Bojke, L., 2020. Modelling the impact of physical activity on public health: a review and critique. Health Policy, 124(10), pp.1155-1164.

Candio, P., Pouwels, K.B., Meads, D., Hill, A.J., Bojke, L. and Williams, C., 2021. Modelling decay in effectiveness for evaluation of behaviour change interventions: a tutorial for public health economists. The European Journal of Health Economics, pp.1-7.

Caro, J.J., Briggs, A.H., Siebert, U. and Kuntz, K.M., 2012. Modeling good research practices—overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Medical Decision Making, 32(5), pp.667-677.

Caspersen, C.J., Powell, K.E. and Christenson, G.M., 1985. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public health reports, 100(2), p.126.

Cavill, N., Kahlmeier, S., Rutter, H., Racioppi, F. and Oja, P., 2008. Economic analyses of transport infrastructure and policies including health effects related to cycling and walking: a systematic review. Transport policy, 15(5), pp.291-304.

Cerin, E. and Leslie, E., 2008. How socio-economic status contributes to participation in leisure-time physical activity. Social science & medicine, 66(12), pp.2596-2609.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2022). shiny: Web Application Framework for R. R package version 1.7.2. [Online]. Available: https://CRAN.R-project.org/package=shiny

Christofaro, D.G.D., Andersen, L.B., Andrade, S.M.D., Barros, M.V.G.D., Saraiva, B.T.C., Fernandes, R.A. and Ritti-Dias, R.M., 2018. Adolescents' physical activity is associated with previous and current physical activity practice by their parents. Jornal de pediatria, 94, pp.48-55.

Church, T.S., Thomas, D.M., Tudor-Locke, C., Katzmarzyk, P.T., Earnest, C.P., Rodarte, R.Q., Martin, C.K., Blair, S.N. and Bouchard, C., 2011. Trends over 5 decades in US occupation-related physical activity and their associations with obesity. PloS one, 6(5), p.e19657.

Chyou, P.H., Scheuer, D. and Linneman, J.G., 2006. Assessment of female participation in an employee 20-week walking incentive program at Marshfield Clinic, a large multispecialty group practice. Clinical Medicine & Research, 4(4), pp.256-265.

Clarke P, Gray A, Legood R, et al. 2003. The impact of diabetes-related complications on healthcare costs: results from the United Kingdom prospective diabetes study (UKPDS study No. 65). Diabet Med; 20:442–50.

Claxton, K., Sculpher, M. and Culyer, A., 2007. Mark versus Luke? Appropriate methods for the evaluation of public health interventions.

Cobiac, L.J., Vos, T. and Barendregt, J.J., 2009. Cost-effectiveness of interventions to promote physical activity: a modelling study. PLoS medicine, 6(7), p.e1000110.

Cochrane, M., Watson, P.M., Timpson, H., Haycox, A., Collins, B., Jones, L., Martin, A. and Graves, L.E.F., 2019. Systematic review of the methods used in economic evaluations of targeted physical activity and sedentary behaviour interventions. Social Science & Medicine, 232, pp.156-167.

Cohen, J.T. and Wong, J.B., 2017. Can economic model transparency improve provider interpretation of cost-effectiveness analysis? A response. Medical care, 55(11), pp.912-914.

Cohen, J.T., Neumann, P.J. and Wong, J.B., 2017. A call for open-source cost-effectiveness analysis. Annals of internal medicine, 167(6), pp.432-433.

Cookson, R., Mirelman, A.J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O.F., Verguet, S. and Culyer, A.J., 2017. Using cost-effectiveness analysis to address health equity concerns. Value in Health, 20(2), pp.206-212.

Craggs, C., Corder, K., van Sluijs, E.M. and Griffin, S.J., 2011. Determinants of change in physical activity in children and adolescents: a systematic review. American journal of preventive medicine, 40(6), pp.645-658.

Craig, C.L., Gauvin, L., Cragg, S., Katzmarzyk, P.T., Stephens, T., Russell, S.J., Bentz, L. and Potvin, L., 2005. Towards a social epidemiological perspective on physical activity and health: the aims, design, and methods of the Physical Activity Longitudinal Study (PALS). Journal of Physical Activity and Health, 2(3), pp.272-284.

Craig, C.L., Marshall, A.L., Sjöström, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U.L.F., Yngve, A., Sallis, J.F. and Oja, P., 2003. International physical activity questionnaire: 12-country reliability and validity. Medicine & science in sports & exercise, 35(8), pp.1381-1395.

Craig, C.L., Russell, S.J., Cameron, C. and Bauman, A., 2004. Twenty-year trends in physical activity among Canadian adults. Canadian journal of public health, 95, pp.59-63.

Craig, R., Mindell, J. and Hirani, V., 2013. Health survey for England. Health and Social Care Information Centre.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P. and Parkin, D., 2015. The influence of cost-effectiveness and other factors on nice decisions. Health economics, 24(10), pp.1256-1271.

Daskalopoulou, C., Stubbs, B., Kralj, C., Koukounari, A., Prince, M. and Prina, A.M., 2017. Physical activity and healthy ageing: A systematic review and meta-analysis of longitudinal cohort studies. Ageing research reviews, 38, pp.6-17.

Davies, S., Burns, H., Jewell, T. and McBride, M., 2011. Start active, stay active: a report on physical activity from the four home countries. Chief Medical Officers, 16306, pp.1-62.

Davis, J.C., Robertson, M.C., Ashe, M.C., Liu-Ambrose, T., Khan, K.M. and Marra, C.A., 2010. Does a home-based strength and balance programme in people aged≥ 80 years provide the best value for money to prevent falls? A systematic review of economic evaluations of falls prevention interventions. British journal of sports medicine, 44(2), pp.80-89.

Deenihan, G. and Caulfield, B., 2014. Estimating the health economic benefits of cycling. Journal of Transport & Health, 1(2), pp.141-149.

Department of Health and Social Care (2019). UK Chief Medical Officers' Physical Activity Guidelines. [online] gov.uk. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/832868/uk-chief-medical-officers-physical-activity-guidelines.pdf. [Accessed 05/02/2023].

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L. and Torrance, G.W., 2015. Methods for the economic evaluation of health care programmes. Oxford university press.

Ding, D., Lawson, K.D., Kolbe-Alexander, T.L., Finkelstein, E.A., Katzmarzyk, P.T., Van Mechelen, W. and Pratt, M., 2016. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. The Lancet, 388(10051), pp.1311-1324.

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M.H., White, T., Van Hees, V.T., Trenell, M.I., Owen, C.G. and Preece, S.J., 2017. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PloS one, 12(2), p.e0169649.

Dumith, S.C., Gigante, D.P., Domingues, M.R. and Kohl III, H.W., 2011a. Physical activity change during adolescence: a systematic review and a pooled analysis. International journal of epidemiology, 40(3), pp.685-698.

Dumith, S.C., Hallal, P.C., Reis, R.S. and Kohl III, H.W., 2011b. Worldwide prevalence of physical inactivity and its association with human development index in 76 countries. Preventive medicine, 53(1-2), pp.24-28.

Dyrstad, S.M., Hansen, B.H., Holme, I.M. and Anderssen, S.A., 2014. Comparison of self-reported versus accelerometer-measured physical activity. Medicine & science in sports & exercise, 46(1), pp.99-106.

Eberhardt, T., Niessner, C., Oriwol, D., Buchal, L., Worth, A. and Bös, K., 2020. Secular trends in physical fitness of children and adolescents: a review of large-scale epidemiological studies published after 2006. International journal of environmental research and public health, 17(16), p.5671.

Eddy, D.M., Hollingworth, W., Caro, J.J., Tsevat, J., McDonald, K.M. and Wong, J.B., 2012. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. Medical Decision Making, 32(5), pp.733-743.

Ekelund, U., Anderssen, S.A., Froberg, K., Sardinha, L.B., Andersen, L.B., Brage, S. and European Youth Heart Study Group, 2007. Independent associations of physical activity and cardiorespiratory fitness with metabolic risk factors in children: the European youth heart study. Diabetologia, 50, pp.1832-1840.

Elvidge, J., Summerfield, A., Nicholls, D. and Dawoud, D., 2022. Diagnostics and treatments of COVID-19: a living systematic review of economic evaluations. Value in Health.

Farrell, L., Hollingsworth, B., Propper, C. and Shields, M.A., 2014. The socioeconomic gradient in physical inactivity: evidence from one million adults in England. Social science & medicine, 123, pp.55-63.

Farooq, A., Martin, A., Janssen, X., Wilson, M.G., Gibson, A.M., Hughes, A. and Reilly, J.J., 2020. Longitudinal changes in moderate-to-vigorous-intensity physical activity in children and adolescents: A systematic review and meta-analysis. Obesity Reviews, 21(1), p.e12953.

Fordham R., Barton, G., 2008. A cost-effectiveness scenario analysis of four interventions to increase child and adolescent physical activity: the case of walking buses, free swimming, dance classes and community sports: NICE Programme Guidance on the Promotion of Physical Activity in Children. [Available from http://www.nice.org.uk/media/C83/74/PromotingPhysicalActivityChildrenCostEffectivenessAnalysis.pdf]

Foster, C., Richards, J., Thorogood, M. and Hillsdon, M., 2013. Remote and web 2.0 interventions for promoting physical activity. Cochrane Database of Systematic Reviews, (9).

Frew, E., Brown, H., Goodman, A., Platt, M., Seidler, E., Spetz, J. and McIntyre, D., 2018. Building an international health economics teaching network. Health Economics.

Frew, E.J., Bhatti, M., Win, K., Sitch, A., Lyon, A., Pallan, M. and 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(3), pp.207-212.

Gao, J., Helbich, M., Dijst, M. and Kamphuis, C.B., 2017. Socioeconomic and demographic differences in walking and cycling in the Netherlands: How do these translate into differences in health benefits?. Journal of Transport & Health, 6, pp.358-365.

Garrett, S., Elley, C.R., Rose, S.B., O'Dea, D., Lawton, B.A. and Dowell, A.C., 2011. Are physical activity interventions in primary care and the community cost-effective? A systematic review of the evidence. British Journal of General Practice, 61(584), pp.e125-e133.

Gc, V.S., Suhrcke, M., Atkin, A.J., Van Sluijs, E. and Turner, D., 2019. Cost-effectiveness of physical activity interventions in adolescents: model development and illustration using two exemplar interventions. BMJ open, 9(8), p.e027566.

Geidl, W., Schlesinger, S., Mino, E., Miranda, L. and Pfeifer, K., 2020. Dose–response relationship between physical activity and mortality in adults with noncommunicable diseases: a systematic review and meta-analysis of prospective observational studies. International Journal of Behavioral Nutrition and Physical Activity, 17(1), pp.1-18.

Genter, J,. Donovan, S., Petrenas, B., Badland, H. 2008 Valuing the health benefits of active travel modes, Aukland: New Zealand Transport Agency Research Report 359.

Giles-Corti, B. and Donovan, R.J., 2002. The relative influence of individual, social and physical environment determinants of physical activity. Social science & medicine, 54(12), pp.1793-1812.

Gordon, L., Graves, N., Hawkes, A. and Eakin, E., 2007. A review of the cost-effectiveness of face-to-face behavioural interventions for smoking, physical activity, diet and alcohol. Chronic illness, 3(2), pp.101-129.

Grandes, G., García-Alvarez, A., Ansorena, M., Sánchez-Pinilla, R.O., Torcal, J., Arietaleanizbeaskoa, M.S. and Sanchez, A., 2023. Any increment in physical activity reduces mortality risk of physically inactive patients: prospective cohort study in primary care. British Journal of General Practice, 73(726), pp.e52-e58.

Griffiths, L.J., Cortina-Borja, M., Sera, F., Pouliou, T., Geraci, M., Rich, C., Cole, T.J., Law, C., Joshi, H., Ness, A.R. and Jebb, S.A., 2013. How active are our children? Findings from the Millennium Cohort Study. BMJ open, 3(8), p.e002893.

Grzywacz, J.G. and Marks, N.F., 2001. Social inequalities and exercise during adulthood: toward an ecological perspective. Journal of Health and Social behavior, pp.202-220.

Gu, J.K., Charles, L.E., Ma, C.C., Andrew, M.E., Fekedulegn, D., Hartley, T.A., Violanti, J.M. and Burchfiel, C.M., 2016. Prevalence and trends of leisure-time physical activity by occupation and industry in US workers: the National Health Interview Survey 2004–2014. Annals of epidemiology, 26(10), pp.685-692.

Gulliford, M.C., Charlton, J., Bhattarai, N., Charlton, C. and Rudisill, C., 2014. Impact and cost-effectiveness of a universal strategy to promote physical activity in primary care: population-based cohort study and Markov model. The European Journal of Health Economics, 15, pp.341-351.

Guthold, R., Stevens, G.A., Riley, L.M. and Bull, F.C., 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1· 9 million participants. The lancet global health, 6(10), pp.e1077-e1086.

Guthold, R., Stevens, G.A., Riley, L.M. and Bull, F.C., 2020. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1· 6 million participants. The Lancet Child & Adolescent Health, 4(1), pp.23-35.

Haake, S., Heller, B., Schneider, P., Smith, R. and Green, G., 2022. The influence of neighbourhood equity on parkrunners in a British city. Health promotion international, 37(2), p.daab138.

Hallal, P.C., Andersen, L.B., Bull, F.C., Guthold, R., Haskell, W. and Ekelund, U., 2012. Global

physical activity levels: surveillance progress, pitfalls, and prospects. The lancet, 380(9838), pp.247-257.

Hart, R., Burns, D., Ramaekers, B., Ren, S., Gladwell, D., Sullivan, W., Davison, N., Saunders, O., Sly, I., Cain, T. and Lee, D., 2020. R and Shiny for cost-effectiveness analyses: why and when? A hypothetical case study. PharmacoEconomics, 38(7), pp.765-776.

Haskell, W.L., Lee, I.M., Pate, R.R., Powell, K.E., Blair, S.N., Franklin, B.A., Macera, C.A., Heath, G.W., Thompson, P.D. and Bauman, A., 2007. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation, 116(9), p.1081.

Hazel, C.A., Bull, S., Greenwell, E., Bunik, M., Puma, J. and Perraillon, M., 2021. Systematic review of cost-effectiveness analysis of behavior change communication apps: Assessment of key methods. Digital Health, 7, p.20552076211000559.

Herrmann, S.D., Heumann, K.J., Der Ananian, C.A. and Ainsworth, B.E., 2013. Validity and reliability of the global physical activity questionnaire (GPAQ). Measurement in Physical Education and Exercise Science, 17(3), pp.221-235.

Hindley, D., 2022. parkrun: An Organised Running Revolution. Taylor & Francis.

Hoffmann, C., Stoykova, B.A., Nixon, J., Glanville, J.M., Misso, K. and Drummond, M.F., 2002. Do health-care decision makers find economic evaluations useful? The findings of focus group research in UK health authorities. Value in Health, 5(2), pp.71-78.

House of Commons (2021). Coronavirus: Lessons Learned to Date Sixth Report of the Health and Social Care Committee and Third Report of the Science and Technology Committee of Session 2021-22 HC 92. [online] Available at: https://committees.parliament.uk/publications/7496/documents/78687/default/.

Hovsepian, S., Kelishadi, R., Motlagh, M.E., Kasaeian, A., Shafiee, G., Arefirad, T., Najafi, F., Khoramdad, M., Asayesh, H., Heshmat, R. and Qorbani, M., 2016. Level of physical activity and screen time among Iranian children and adolescents at the national and provincial level: The CASPIAN-IV study. Medical journal of the Islamic Republic of Iran, 30, p.422.

HSE. Health Survey for England. 2004. Available from https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2004-health-of-ethnic-minorities-headline-results

HSE. Health Survey for England. 2008. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2008-physical-activity-and-fitness

HSE. Health Survey for England. 2014. Available from https://digital.nhs.uk/data-and-information/publications/statistical/health-survey-for-england/health-survey-for-england-2014

Hu, G., Jousilahti, P., Barengo, N.C., Qiao, Q., Lakka, T.A. and Tuomilehto, J., 2005. Physical activity, cardiovascular risk factors, and mortality among Finnish adults with diabetes. Diabetes care, 28(4), pp.799-805.

Hu, G., Jousilahti, P., Borodulin, K., Barengo, N.C., Lakka, T.A., Nissinen, A. and Tuomilehto, J., 2007. Occupational, commuting and leisure-time physical activity in relation to coronary heart disease among middle-aged Finnish men and women. Atherosclerosis, 194(2), pp.490-497.

Hu, G., Qiao, Q., Silventoinen, K., Eriksson, J.G., Jousilahti, P., Lindström, J., Valle, T.T., Nissinen, A. and Tuomilehto, J., 2003. Occupational, commuting, and leisure-time physical activity in relation to risk for Type 2 diabetes in middle-aged Finnish men and women. Diabetologia, 46, pp.322-329.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A.H., Mauskopf, J., Loder, E. and ISPOR Health Economic Evaluation Publication

Guidelines-CHEERS Good Reporting Practices Task Force, 2013. Consolidated health economic evaluation reporting standards (CHEERS)—explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value in health, 16(2), pp.231-250.

Incerti, D., Thom, H., Baio, G. and Jansen, J.P., 2019. R you still using excel? The advantages of modern software tools for health technology assessment. Value in Health, 22(5), pp.575-579.

Ishii, K., Shibata, A. and Oka, K., 2010. Environmental, psychological, and social influences on physical activity among Japanese adults: structural equation modeling analysis. International Journal of Behavioral Nutrition and Physical Activity, 7(1), pp.1-8.

Jackson, C.H., Thompson, S.G. and Sharples, L.D., 2009. Accounting for uncertainty in health economic decision models by using model averaging. Journal of the Royal Statistical Society: Series A (Statistics in Society), 172(2), pp.383-404.

Jago, R., Sebire, S.J., Davies, B., Wood, L., Edwards, M.J., Banfield, K., Fox, K.R., Thompson, J.L., Powell, J.E. and Montgomery, A.A., 2014. Randomised feasibility trial of a teaching assistant led extracurricular physical activity intervention for 9 to 11 year olds: Action 3: 30. International Journal of Behavioral Nutrition and Physical Activity, 11(1), pp.1-14.

Jalal, H., Pechlivanoglou, P., Krijkamp, E., Alarid-Escudero, F., Enns, E. and Hunink, M.M., 2017. An overview of R in health decision sciences. Medical decision making, 37(7), pp.735-746.

Janssen, I. and LeBlanc, A.G., 2010. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. International journal of behavioral nutrition and physical activity, 7(1), pp.1-16.

Johri, M. and Norheim, O.F., 2012. Can cost-effectiveness analysis integrate concerns for equity? Systematic review. International journal of technology assessment in health care, 28(2), pp.125-132.

Kaambwa B, Frew E. 2013. Health economics in the UK: capacity, constraints and comparisons to the US health economists. International Reviews of Economics Education 2013;12:1–11.

Kahlmeier, S., Götschi, T., Cavill, N., Castro Fernandez, A., Brand, C., Rojas Rueda, D., Woodcock, J., Kelly, P., Lieb, C., Oja, P. and Foster, C., 2017. Health economic assessment tool (HEAT) for walking and for cycling: Methods and user guide on physical activity, air pollution, injuries and carbon impact assessments.

Kahlmeier, S., Racioppi, F., Götschi, T., Castro, A. and Cavill, N., 2020. The WHO health economic assessment tool for walking and cycling: How to quantify impacts of active mobility. In Advances in Transportation and Health (pp. 329-342). Elsevier.

Katzmarzyk, P.T., Lee, I.M., Martin, C.K. and Blair, S.N., 2017. Epidemiology of physical activity and exercise training in the United States. Progress in cardiovascular diseases, 60(1), pp.3-10.

Kelly, M.P., McDaid, D., Ludbrook, A. and Powell, J., 2005. Economic appraisal of public health interventions. London: Health Development Agency.

Kent, S., Becker, F., Feenstra, T., Tran-Duy, A., Schlackow, I., Tew, M., Zhang, P., Ye, W., Lizheng, S., Herman, W. and McEwan, P., 2019. The challenge of transparency and validation in health economic decision modelling: a view from Mount Hood. Pharmacoeconomics, 37(11), pp.1305-1312.

Kingham S, Ussher S, Ticket to a sustainable future: An evaluation of the longterm durability of the Walking School Bus programme in Christchurch, New Zealand. Transport Policy 2005;12:314-323.

Kohrt, W.M., Bloomfield, S.A., Little, K.D., Nelson, M.E. and Yingling, V.R., 2004. Physical activity and bone health. Medicine & Science in Sports & Exercise, 36(11), pp.1985-1996.

Kyu, H.H., Bachman, V.F., Alexander, L.T., Mumford, J.E., Afshin, A., Estep, K., Veerman, J.L., Delwiche, K., Iannarone, M.L., Moyer, M.L. and Cercy, K., 2016. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic

review and dose-response meta-analysis for the Global Burden of Disease Study 2013. bmj, 354.

Laine, J., Kuvaja-Köllner, V., Pietilä, E., Koivuneva, M., Valtonen, H. and Kankaanpää, E., 2014. Cost-effectiveness of population-level physical activity interventions: a systematic review. American journal of health promotion, 29(2), pp.71-80.

Lee, I.M. and Shiroma, E.J., 2014. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. British journal of sports medicine, 48(3), pp.197-201.

Leeds University Libraries. 2021. The greatest possible impact. Webpage available at: https://leedsunilibrary.wordpress.com/2021/06/25/the-greatest-possible-impact-the-wellcome-trust-and-open-research/

Lehnert, T., Sonntag, D., Konnopka, A., Riedel-Heller, S. and König, H.H., 2012. The long-term cost-effectiveness of obesity prevention interventions: systematic literature review. Obesity reviews, 13(6), pp.537-553.

Lewis, B.A., Marcus, B.H., Pate, R.R. and Dunn, A.L., 2002. Psychosocial mediators of physical activity behavior among adults and children. American journal of preventive medicine, 23(2), pp.26-35.

Lewis, C., Ubido, J., Holford, R. and Scott-Samuel, A., 2010. Prevention programmes cost-effectiveness review. Liverpool Public Health Observatory Report,(https://www. liverpool. ac. uk/media/livacuk/instituteofpsychology/researchgroups/lpho/83_28th_Feb_Physical_activity_and_cost_FINAL. pdf. Accessed 06 Apr 2017).

Lounassalo, I., Salin, K., Kankaanpää, A., Hirvensalo, M., Palomäki, S., Tolvanen, A., Yang, X. and Tammelin, T.H., 2019. Distinct trajectories of physical activity and related factors during the life course in the general population: a systematic review. BMC Public Health, 19, pp.1-12.

Love-Koh, J., Taylor, M. 2017. National Institute for Health and Care Excellence. Physical Activity and the Environment: Final Report. Available from https://www.nice.org.uk/guidance/ng90/evidence/economic-modelling-report-pdf-4788819757.

Loyen, A., Clarke-Cornwell, A.M., Anderssen, S.A., Hagströmer, M., Sardinha, L.B., Sundquist, K., Ekelund, U., Steene-Johannessen, J., Baptista, F., Hansen, B.H. and Wijndaele, K., 2016. Sedentary time and physical activity surveillance through accelerometer pooling in four European countries. Sports Medicine, pp.1-15.

Lozano, C. et al. (2012) "Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010." Lancet, 380(9859), pp. 2095-2128.

Madan J, Rawdin A, Stevenson M, et al. A rapid-response economic evaluation of the UK NHS cancer reform strategy breast cancer screening program extension via a plausible bounds approach. Value Health 2010;13:215–21.

Mammen, G. and Faulkner, G., 2013. Physical activity and the prevention of depression: a systematic review of prospective studies. American journal of preventive medicine, 45(5), pp.649-657.

Marks, L., Weatherly, H. and Mason, A., 2013. Prioritizing investment in public health and health equity: what can commissioners do?. Public Health, 127(5), pp.410-418.

McCabe, C. and Dixon, S., 2000. Testing the validity of cost-effectiveness models. *Pharmacoeconomics*, *17*, pp.501-513.

Melzer, K., Kayser, B. and Pichard, C., 2004. Physical activity: the health benefits outweigh the risks. Current Opinion in Clinical Nutrition & Metabolic Care, 7(6), pp.641-647.

Mears, R. and Jago, R., 2016. Effectiveness of after-school interventions at increasing moderate-to-vigorous physical activity levels in 5-to 18-year olds: a systematic review and meta-analysis. British Journal of Sports Medicine, 50(21), pp.1315-1324.

Minton, J., Dimairo, M., Everson-Hock, E., Scott, E. and Goyder, E., 2013. Exploring the relationship between baseline physical activity levels and mortality reduction associated with increases in physical activity: a modelling study. BMJ open, 3(10), p.e003509.

Moore, S.C., Lee, I.M., Weiderpass, E., Campbell, P.T., Sampson, J.N., Kitahara, C.M., Keadle, S.K., Arem, H., De Gonzalez, A.B., Hartge, P. and Adami, H.O., 2016. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA internal medicine, 176(6), pp.816-825.

Morris, J.N., Heady, J.A., Raffle, P.A.B., Roberts, C.G. and Parks, J.W., 1953. Coronary heart-disease and physical activity of work. The lancet, 262(6796), pp.1111-1120.

Müller-Riemenschneider, F., Reinhold, T. and Willich, S.N., 2009. Cost-effectiveness of interventions promoting physical activity. British journal of sports medicine, 43(1), pp.70-76.

Mytton, O.T., Tainio, M., Ogilvie, D., Panter, J., Cobiac, L. and Woodcock, J., 2017. The modelled impact of increases in physical activity: the effect of both increased survival and reduced incidence of disease. European Journal of Epidemiology, 32, pp.235-250.

Naylor, N.R., Williams, J., Green, N., Lamrock, F. and Briggs, A., 2022. Extensions of Health Economic Evaluations in R for Microsoft Excel Users: A Tutorial for Incorporating Heterogeneity and Conducting Value of Information Analyses. PharmacoEconomics, pp.1-12.

NICE. 2014. Developing NICE guidelines: the manual Process and methods. [Online]Available: www.nice.org.uk/process/pmg20

NICE. 2007. Intervention Review: Children and Active Travel. Public Health Collaborating Centre – Physical activity 29th November 2007 (Revised July 2008)

NICE. 2009. Physical activity for children and young people. Public Health Guidance. Available from: https://www.nice.org.uk/guidance/ph17/resources/physical-activity-for-children-and-young-people-pdf-1996181580229

O'Mahony, J.F., Newall, A.T. and van Rosmalen, J., 2015. Dealing with time in health economic evaluation: methodological issues and recommendations for practice. Pharmacoeconomics, 33(12), pp.1255-1268.

Ortega, F.B., Konstabel, K., Pasquali, E., Ruiz, J.R., Hurtig-Wennlöf, A., Mäestu, J., Löf, M., Harro, J., Bellocco, R., Labayen, I. and Veidebaum, T., 2013. Objectively measured physical activity and sedentary time during childhood, adolescence and young adulthood: a cohort study. PloS one, 8(4), p.e60871.

Østerås, H., & Hammer, S. (2006). The effectiveness of a pragmatic worksite physical activity program on maximal oxygen consumption and the physical activity level in healthy people. Journal of Bodywork and Movement Therapies, 10, 51-57.

Owen, R.K., Bradbury, N., Xin, Y., Cooper, N. and Sutton, A., 2019. MetaInsight: an interactive web-based tool for analyzing, interrogating, and visualizing network meta-analyses using R-shiny and netmeta. Research synthesis methods, 10(4), pp.569-581.

Pareto, V. 1906. Manual of Political Economy. New York, NY: Macmillan.

Parkrun Blog. 2020. Using research to improve inclusivity. Available at: https://blog.parkrun.com/uk/2020/12/08/using-research-to-improve-inclusivity/

Pavey, T.G., Taylor, A.H., Fox, K.R., Hillsdon, M., Anokye, N., Campbell, J.L., Foster, C., Green, C., Moxham, T., Mutrie, N. and Searle, J., 2011. Effect of exercise referral schemes in primary care on physical activity and improving health outcomes: systematic review and meta-analysis. Bmj, 343.

Penedo, F.J. and Dahn, J.R., 2005. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Current opinion in psychiatry, 18(2), pp.189-193.

Peters, H., Whincup, P.H., Cook, D.G., Law, C. and Li, L., 2014. Trends in resting pulse rates in 9–

11-year-old children in the UK 1980–2008. Archives of disease in childhood, 99(1), pp.10-14.

Plasqui, G., Bonomi, A.G. and Westerterp, K.R., 2013. Daily physical activity assessment with accelerometers: new insights and validation studies. obesity reviews, 14(6), pp.451-462.

Pollitt, R.A., Rose, K.M. and Kaufman, J.S., 2005. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. BMC public health, 5, pp.1-13.

Poulton, R., Caspi, A., Milne, B.J., Thomson, W.M., Taylor, A., Sears, M.R. and Moffitt, T.E., 2002. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study. The lancet, 360(9346), pp.1640-1645.

Pouwels, X.G., Sampson, C.J., Arnold, R.J., Janodia, M.D., Henderson, R., Lamotte, M., Cowell, W., Borrill, J., Huttin, C., Udupa, N. and Gong, C.L., 2022. Opportunities and Barriers to the Development and Use of Open Source Health Economic Models: A Survey. Value in health, 25(4), pp.473-479.

Purath, J., Miller, A.M., McCabe, G. and Wilbur, J., 2004. A brief intervention to increase physical activity in sedentary working women. Canadian Journal of Nursing Research Archive, pp.76-91.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rebar, A.L., Stanton, R., Geard, D., Short, C., Duncan, M.J. and Vandelanotte, C., 2015. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health psychology review, 9(3), pp.366-378.

Reece, L.J., Quirk, H., Wellington, C., Haake, S.J. and Wilson, F., 2019. Bright Spots, physical activity investments that work: Parkrun; a global initiative striving for healthier and happier communities. British journal of sports medicine, 53(6), pp.326-327.

Reiner, M., Niermann, C., Jekauc, D. and Woll, A., 2013. Long-term health benefits of physical activity–a systematic review of longitudinal studies. BMC public health, 13(1), pp.1-9.

Rhodes, R.E. and Nigg, C.R., 2011. Advancing physical activity theory: A review and future directions. Exercise and sport sciences reviews, 39(3), pp.113-119.

Ridley, K. and Olds, T., 2008. Assigning energy costs to activities in children: a review and synthesis. Medicine+ Science in Sports+ Exercise, 40(8), p.1439.

Rütten, A., Vuillemin, A., Ooijendijk, W.T.M., Schena, F., Sjöström, M., Stahl, T., Auweele, Y.V., Welshman, J. and Ziemainz, H., 2003. Physical activity monitoring in europe. The European Physical Activity Surveillance System (EUPASS) approach and indicator testing. Public health nutrition, 6(4), pp.377-384.

Sallis, J.F., Cerin, E., Conway, T.L., Adams, M.A., Frank, L.D., Pratt, M., Salvo, D., Schipperijn, J., Smith, G., Cain, K.L. and Davey, R., 2016. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. The lancet, 387(10034), pp.2207-2217.

Sallis, J.F., Prochaska, J.J. and Taylor, W.C., 2000. A review of correlates of physical activity of children and adolescents. Medicine and science in sports and exercise, 32(5), pp.963-975.

Sampson, C.J., Arnold, R., Bryan, S., Clarke, P., Ekins, S., Hatswell, A., Hawkins, N., Langham, S., Marshall, D., Sadatsafavi, M. and Sullivan, W., 2019. Transparency in decision modelling: what, why, who and how?. Pharmacoeconomics, 37(11), pp.1355-1369.

Sanders, S., 2017. National life tables, UK: 2014 to 2016.

Santtila, M., Kyröläinen, H., Vasankari, T., Tiainen, S., Palvalin, K., Häkkinen, A. and Häkkinen, K., 2006. Physical fitness profiles in young Finnish men during the years 1975-2004. Medicine & Science in Sports & Exercise, 38(11), pp.1990-1994.

Schneider, P.P., Smith, R.A., Bullas, A.M., Quirk, H., Bayley, T., Haake, S.J., Brennan, A. and

Goyder, E., 2020. Multiple deprivation and geographic distance to community physical activity events—Achieving equitable access to parkrun in England. Public health, 189, pp.48-53.

Shields, G.E., Pennington, B., Bullement, A., Wright, S. and Elvidge, J., 2022. Out of Date or Best Before? A Commentary on the Relevance of Economic Evaluations Over Time. PharmacoEconomics, 40(3), pp.249-256.

Scholes, S. and Mindell, J., 2016. Health survey for England 2015 physical activity in children. *Health and Social Care Information Centre*.

Siemieniuk R A, Bartoszko J J, Zeraatkar D, Kum E, Qasim A, DÃaz Martinez J P et al. Drug treatments for covid-19: living systematic review and network meta-analysis BMJ 2020; 370 :m2980 doi:10.1136/bmj.m2980

Sloman 2009. Analysis and Synthesis of Evidence on the Effects of Investment in Six Cycling Demonstration Towns http://www.ciltuk.org.uk/download/Analysis_and_synthesis_of_evidence_on_the_effects_of_investment.pdf

Smith, R.D. and Petticrew, M., 2010. Public health evaluation in the twenty-first century: time to see the wood as well as the trees. Journal of Public Health, 32(1), pp.2-7.

Smith, R. and Schneider, P., 2020a. Making health economic models Shiny: A tutorial. Wellcome open research, 5.

Smith, R., Schneider, P., Bullas, A., Haake, S., Quirk, H., Cosulich, R. and Goyder, E., 2020b. Does ethnic density influence community participation in mass participation physical activity events? The case of parkrun in England. Wellcome Open Research, 5.

Smith, R., Thomas, C., Squires, H., Götschi, T., Kahlmeier, S. and Goyder, E., 2021a. The price of precision: trade-offs between usability and validity in the World Health Organization Health Economic Assessment Tool for walking and cycling. Public Health, 194, pp.263-269.

Smith, R.A., Schneider, P.P., Cosulich, R., Quirk, H., Bullas, A.M., Haake, S.J. and Goyder, E., 2021b. Socioeconomic inequalities in distance to and participation in a community-based running and walking activity: A longitudinal ecological study of parkrun 2010 to 2019. Health & Place, 71, p.102626.

Smith, R. and Schneider, P. 2021c. Dark-peak-analytics/darkpeak, GitHub. Available at: https://github.com/dark-peak-analytics/darkpeak (Accessed: January 26, 2023).

Smith, R.A., Schneider, P.P. and Mohammed, W., 2022a. Living HTA: automating health technology assessment with R. Wellcome Open Research, 7, p.194.

Smith, J.L., Reece, L.J., Rose, C.L. and Owen, K.B., 2022b. Improving access to public physical activity events for disadvantaged communities in Australia. BMC Public Health, 22(1), pp.1-8.

Smith, R., Thomas, C., Squires, H. and Goyder, E., 2022c. A comparison of the World Health Organisation's HEAT model results using a non-linear physical activity dose response function with results from the existing tool. Wellcome Open Research, 7.

Spence, J.C. and Lee, R.E., 2003. Toward a comprehensive model of physical activity. Psychology of sport and exercise, 4(1), pp.7-24.

Sport England. 2018. Sport England Partner with Parkrun for 3 years with 3million investment. Available at: https://www.sportengland.org/news/sport-england-partner-with-parkrun-for-three-years-with-3-million-investment#:~:text=A%20%C2%A33%20million%20investment,200%20new%20events%20across%20England.

Squires, H., Chilcott, J., Akehurst, R., Burr, J. and Kelly, M.P., 2016a. A systematic literature review of the key challenges for developing the structure of public health economic models. International

journal of public health, 61, pp.289-298.

Squires, H., Chilcott, J., Akehurst, R., Burr, J. and Kelly, M.P., 2016b. A framework for developing the structure of public health economic models. Value in Health, 19(5), pp.588-601.

Stamatakis, E., Ekelund, U. and Wareham, N.J., 2007. Temporal trends in physical activity in England: the Health Survey for England 1991 to 2004. Preventive medicine, 45(6), pp.416-423.

Strong, M., Brennan, A. and Oakley, J., 2015. How to calculate value of information in seconds using 'SAVI', the Sheffield Accelerated Value of Information web app. Value in Health, 18(7), pp.A725-A726.

Sullivan, P.W., Slejko, J.F., Sculpher, M.J. and Ghushchyan, V., 2011. Catalogue of EQ-5D scores for the United Kingdom. Medical Decision Making, 31(6), pp.800-804.

Sun, F., Norman, I.J. and While, A.E., 2013. Physical activity in older people: a systematic review. BMC public health, 13(1), pp.1-17.

Sustrans (2014), Walking and cycling outcomes – Sustran's Scottish Government Grant 2012-2015: Second Interim Report; Sustrans Scotland

Tappenden, P., Brazier, J., Ratcliffe, J. and Chilcott, J., 2007. A stated preference binary choice experiment to explore NICE decision making. Pharmacoeconomics, 25(8), pp.685-693.

Tappenden P, Chilcott J, Eggington S, et al. Option appraisal of population-based colorectal cancer screening programmes in England. Gut 2007;56:677–84.

Thokala, P., Srivastava, T., Smith, R., Ren, S., Whittington, M.D., Elvidge, J., Wong, R. and Uttley, L., 2023. Living Health Technology Assessment: Issues, Challenges and Opportunities. PharmacoEconomics, pp.1-11.

Tian, Y., Jiang, C., Wang, M., Cai, R., Zhang, Y., He, Z., Wang, H., Wu, D., Wang, F., Liu, X. and He, Z., 2016. BMI, leisure-time physical activity, and physical fitness in adults in China: results from a series of national surveys, 2000–14. The lancet Diabetes & endocrinology, 4(6), pp.487-497.

Troiano, R.P., McClain, J.J., Brychta, R.J. and Chen, K.Y., 2014. Evolution of accelerometer methods for physical activity research. British journal of sports medicine, 48(13), pp.1019-1023.

Trost, S.G., Owen, N., Bauman, A.E., Sallis, J.F. and Brown, W., 2002. Correlates of adults' participation in physical activity: review and update. Medicine & science in sports & exercise, 34(12), pp.1996-2001.

Van Dongen, J.M., Proper, K.I., Van Wier, M.F., Van der Beek, A.J., Bongers, P.M., Van Mechelen, W. and Van Tulder, M.W., 2011. Systematic review on the financial return of worksite health promotion programmes aimed at improving nutrition and/or increasing physical activity. Obesity reviews, 12(12), pp.1031-1049.

van Sluijs, E.M., Ekelund, U., Crochemore-Silva, I., Guthold, R., Ha, A., Lubans, D., Oyeyemi, A.L., Ding, D. and Katzmarzyk, P.T., 2021. Physical activity behaviours in adolescence: current evidence and opportunities for intervention. The Lancet, 398(10298), pp.429-442.

Varma, V.R., Dey, D., Leroux, A., Di, J., Urbanek, J., Xiao, L. and Zipunnikov, V., 2017. Re-evaluating the effect of age on physical activity over the lifespan. Preventive medicine, 101, pp.102-108.

Vidoni, M. 2022. Understanding roxygen package documentation in R. Journal of Systems and Software 188: 111265.

Vijay, G.C., Wilson, E.C., Suhrcke, M., Hardeman, W. and Sutton, S., 2016. Are brief interventions to increase physical activity cost-effective? A systematic review. British journal of sports medicine, 50(7), pp.408-417.

Viscusi, W.K. and Aldy, J.E., 2003. The value of a statistical life: a critical review of market estimates throughout the world. Journal of risk and uncertainty, 27(1), pp.5-76.

Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., Abdelalim, A. and Abdollahi, M., 2020. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet, 396(10258), pp.1204-1222.

Warburton, D.E. and Bredin, S.S., 2017. Health benefits of physical activity: a systematic review of current systematic reviews. Current opinion in cardiology, 32(5), pp.541-556.

Ward S, Lloyd Jones M, Pandor A, et al. A systematic review and economic evaluation of statins for the prevention of coronary events. Health Technol Assess 2007;11:1–160.

Weatherly, H., Drummond, M., Claxton, K., Cookson, R., Ferguson, B., Godfrey, C., Rice, N., Sculpher, M. and Sowden, A., 2009. Methods for assessing the cost-effectiveness of public health interventions: key challenges and recommendations. Health policy, 93(2-3), pp.85-92.

Westerterp, K.R., Saris, W.H., Van Es, M. and Ten Hoor, F., 1986. Use of the doubly labeled water technique in humans during heavy sustained exercise. Journal of Applied Physiology, 61(6), pp.2162-2167.

Whitty, C.J., 2015. What makes an academic paper useful for health policy?. BMC medicine, 13(1), pp.1-5.

WHO FGM Cost calculator (2022). World Health Organisation. Available at: https://srhr.org/fgmcost/cost-calculator/ (Accessed: January 12, 2023).

Williams, C.J., Wilson, K.J. and Wilson, N., 2021. A comparison of prior elicitation aggregation using the classical method and SHELF. Journal of the Royal Statistical Society: Series A (Statistics in Society), 184(3), pp.920-940.

Williams, N.H., Hendry, M., France, B., Lewis, R. and Wilkinson, C., 2007. Effectiveness of exercise-referral schemes to promote physical activity in adults: systematic review. British journal of general practice, 57(545), pp.979-986.

Williamson, E.J., Walker, A.J., Bhaskaran, K., Bacon, S., Bates, C., Morton, C.E., Curtis, H.J., Mehrkar, A., Evans, D., Inglesby, P. and Cockburn, J., 2020. Factors associated with COVID-19-related death using OpenSAFELY. Nature, 584(7821), pp.430-436.

Windle, G., Hughes, D., Linck, P., Morgan, R., Burholt, V., Edwards, R.T., Reeves, C., Yeo, S.T., Woods, R. and Russell, I., 2007. Public health interventions to promote mental well-being in people aged 65 and over: systematic review of effectiveness and cost-effectiveness. Bangor: University of Wales.

Windle, G., Hughes, D., Linck, P., Russell, I. and Woods, B., 2010. Is exercise effective in promoting mental well-being in older age? A systematic review. Aging & mental health, 14(6), pp.652-669.

Wolfenstetter, S.B. and Wenig, C.M., 2010. Economic evaluation and transferability of physical activity programmes in primary prevention: a systematic review. International journal of environmental research and public health, 7(4), pp.1622-1648.

World Health Organization Southeast Asia Region. 2023. Physical Activity. Retrieved February 5, 2023, from https://www.who.int/southeastasia/health-topics/physical-activity

World Health Organization, 2004. Global strategy on diet, physical activity and health.

World Health Organization, 2013. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization.

World Health Organization, 2021. Physical activity fact sheet (No. WHO/HEP/HPR/RUN/2021.2). World Health Organization.

Wu, S., Cohen, D., Shi, Y., Pearson, M. and Sturm, R., 2011. Economic analysis of physical activity interventions. American journal of preventive medicine, 40(2), pp.149-158.

Xin, Y., Gray, E., Robles-Zurita, J.A., Haghpanahan, H., Heggie, R., Kohli-Lynch, C., Briggs, A.,

McAllister, D.A., Lawson, K.D. and Lewsey, J., 2022. From spreadsheets to script: experiences from converting a Scottish cardiovascular disease policy model into R. Applied health economics and health policy, 20(2), pp.149-158.

Yang, G., Kong, L., Zhao, W., Wan, X., Zhai, Y., Chen, L.C. and Koplan, J.P., 2008. Emergence of chronic non-communicable diseases in China. The Lancet, 372(9650), pp.1697-1705.

Zwolinsky, S., McKenna, J., Pringle, A., Widdop, P. and Griffiths, C., 2015. Physical activity assessment for public health: efficacious use of the single-item measure. Public Health, 129(12), pp.1630-1636.

Appendix

A1. Literature review extraction form

The completed form can be acquired in xlsx format from the author upon request. The structure of the extraction form is as follows:

Rows:

Section/item, Reference, Title, Abstract, Introduction, Background and objectives, Methods, Target population and subgroups, Setting and location, Study perspective, Comparators, Time horizon, Discount rate, Choice of health outcomes, Measurement of effectiveness, Measurement and valuation of preference based outcomes, Estimating resources and costs, Currency, price date, and conversion, Choice of model, Assumptions, Analytical methods, Results, Study parameters, Incremental costs and outcomes, Characterising uncertainty, Characterising heterogeneity, Discussion, Study findings, limitations, generalisability, and current knowledge, Other, Source of funding, Conflicts of interest, Nature of PA, Inequalities, Lifecourse, Base Population, Explicitly state poor quality evidence Other Sectors included, ICER

Columns:

Section/item, Item No, Recommendation, Love-Koh & Taylor (2017), Campbell et al. (2013), Anokye et al. (2013), Brennan et al. (2012), Fordham & Barton (2008), Bending et al. (2008)

A2. CHEERS 2022 Checklist

Topic	No.	Item	Location where item is reported
Title	1	Identify the study as an economic evaluation and specify the interventions being compared.	Title of section and prelude to chapter
Abstract	2	Provide a structured summary that highlights context, key methods, results, and alternative analyses.	Abstract Section
Introduction			
Background and objectives	3	Give the context for the study, the study question, and its practical relevance for decision making in policy or practice.	Introduction
Methods			
Health economic analysis plan	4	Indicate whether a health economic analysis plan was developed and where available.	Not undertaken
Study population	5	Describe characteristics of the study population (such as age range, demographics, socioeconomic, or clinical characteristics).	Methods, Introduction
Setting and location	6	Provide relevant contextual information that may influence findings.	Introduction
Comparators	7	Describe the interventions or strategies being compared and why chosen.	Methods, The Intervention
Perspective	8	State the perspective(s) adopted by the study and why chosen.	Methods, Scope
Time horizon	9	State the time horizon for the study and why appropriate.	Methods, Scope
Discount rate	10	Report the discount rate(s) and reason chosen.	Methods, Economic Module
Selection of outcomes	11	Describe what outcomes were used as the measure(s) of benefit(s) and harm(s).	Methods, Epidemiological Module
Measurement of outcomes	12	Describe how outcomes used to capture benefit(s) and harm(s) were measured.	Methods, Epidemiological Module
Valuation of outcomes	13	Describe the population and methods used to measure and value outcomes.	Methods, Scope
Measurement and valuation of resources and costs	14	Describe how costs were valued.	Methods, Economic Module
Currency, price date, and conversion	15	Report the dates of the estimated resource quantities and unit costs, plus the currency and year of conversion.	Methods, Economic Module

Rationale and description of model	16	If modelling is used, describe in detail and why used. Report if the model is publicly available and where it can be accessed.	Methods- Model Structure & Methods- Model Code
Analytics and assumptions	17	Describe any methods for analysing or statistically transforming data, any extrapolation methods, and approaches for validating any model used.	Methods
Characterising heterogeneity	18	Describe any methods used for estimating how the results of the study vary for subgroups.	NA
Characterising distributional effects	19	Describe how impacts are distributed across different individuals or adjustments made to reflect priority populations.	NA
Characterising uncertainty	20	Describe methods to characterise any sources of uncertainty in the analysis.	See Methods - Incorporating Probabilistic … and Sensitivity Analysis
Approach to engagement with patients and others affected by the study	21	Describe any approaches to engage patients or service recipients, the general public, communities, or stakeholders (such as clinicians or payers) in the design of the study.	Methods - The model scope
Results			
Study parameters	22	Report all analytic inputs (such as values, ranges, references) including uncertainty or distributional assumptions.	See Methods - The model structure
Summary of main results	23	Report the mean values for the main categories of costs and outcomes of interest and summarise them in the most appropriate overall measure.	Results - Model Results
Effect of uncertainty	24	Describe how uncertainty about analytic judgments, inputs, or projections affect findings. Report the effect of choice of discount rate and time horizon, if applicable.	Results - Sensitivity Analysis
Effect of engagement with patients and others affected by the study	25	Report on any difference patient/service recipient, general public, community, or stakeholder involvement made to the approach or findings of the study	Methods - Introduction paragraph
Discussion			
Study findings, limitations, generalisability, and current knowledge	26	Report key findings, limitations, ethical or equity considerations not captured, and how these could affect patients, policy, or practice.	Discussion
Other relevant information			
Source of funding	27	Describe how the study was funded and any role of the funder in the identification, design, conduct, and reporting of the analysis	NA
Conflicts of interest	28	Report authors conflicts of interest according to journal or International Committee of Medical Journal Editors requirements.	NA

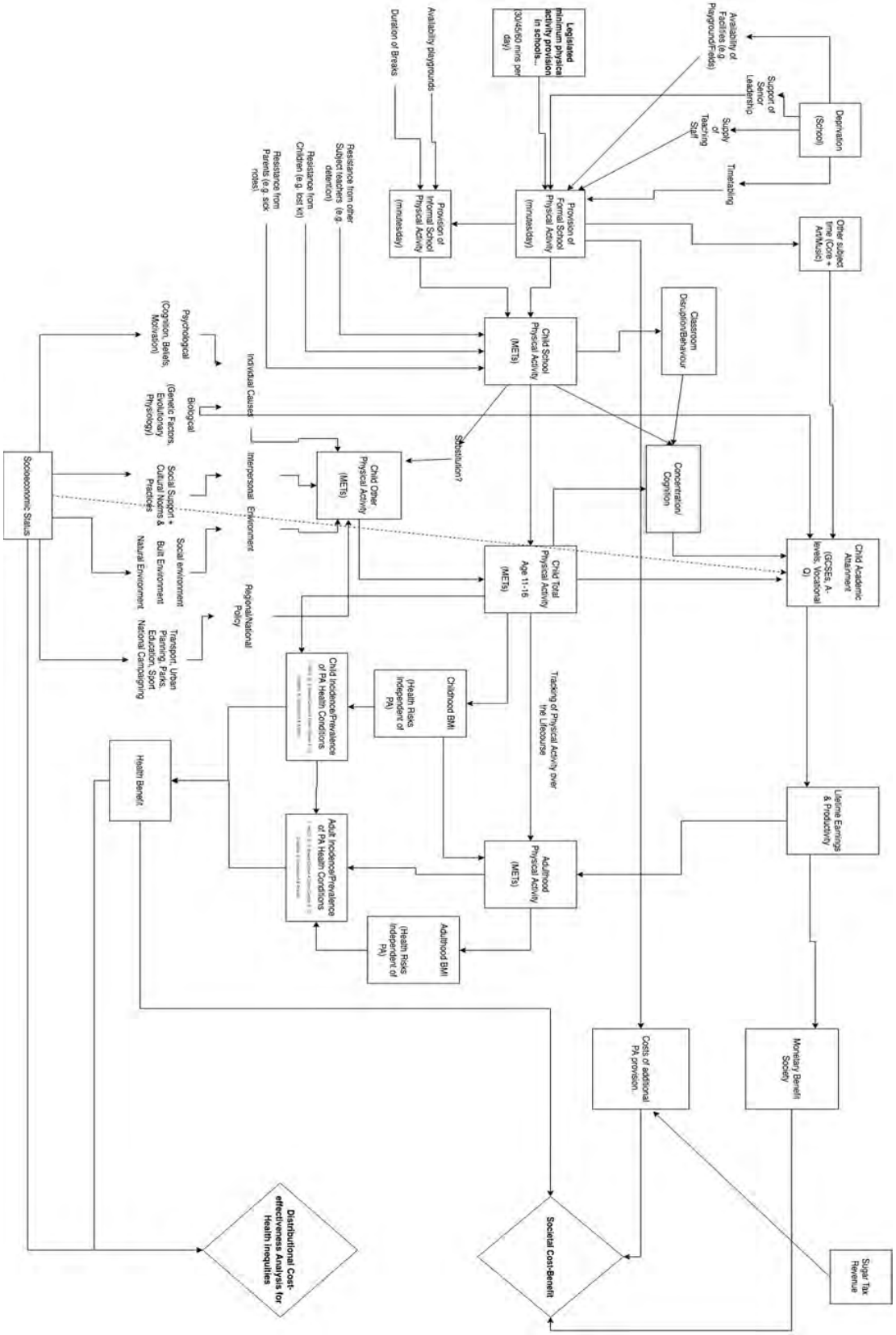
From: Husereau D, Drummond M, Augustovski F, et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. Value Health 2022;25. doi:10.1016/j.jval.2021.10.008

A3. METs for HSE physical activity classifications

The following HSE physical activity classifications (for example nspatT6) were manually compared to the different activities included in Butte et al. 2008 and a metabolic equivalent of task (MET) assigned. These METs are used in the model to estimate total weekly MET minutes for each individual in the model.

HSE Activity code	Estimated METs	HSE Activity code	Estimated METs
nspatT6	4	gymtot08	2.7
nspatT7	4	wkouttot08	2.9
nspatT8	4	aertot08	4.3
nspatT9	4	tentot08	6.4
nspatT10	4	schMonMVPA	6
wepat3	4	schTueMVPA	6
wepat4	4	schWedMVPA	6
cyctot08	5.5	schThurMVPA	6
hoovtot08	3.9	schFriMVPA	6
hoptot08	6.6	schSatMVPA	6
tramtot08	7.2	schSunMVPA	6
playtot08	6	WlkScWT	4
skatot08	5.2	CycScWT	5.5
danctot08	4.3		
skptot08	7.1		
fbltot08	8.2		
nbltot08	5.7		
crkttot08	4.1		
runtot08	8.8		
swmltot08	9		
swmstot08	2.7		
gymtot08	2.7		

A4. Original problem oriented conceptual model



A5. Stability of results in Probabilistic Sensitivity Analysis

Figure showing the stability of the Incremental Cost-Effectiveness Ratio (ICER) on the y-axis and the number of probabilistic sensitivity analysis (PSA) iterations on the x-axis ranging from 500 to 5000. The ICER estimates for the first 500 iterations are excluded due to high expected variance. The dotted red horizontal line represents the mean ICER estimate at 5000 iterations. The shaded ribbon around the mean line represents the range of +/- 5% of the mean ICER value, illustrating the degree of variation in the ICER estimates.