The role of economic modelling in informing the allocation of scarce resources through health technology and health research impact assessment: a critical review

A thesis submitted for the
degree of Doctor of Philosophy
by Published Works

by

Matthew Jonathan Glover

Health Economics Research Group
Department of Health Sciences, Brunel University London

28th February 2022

Abstract

Over the last 30 years, economic evaluation has increasingly been used as a tool to inform the allocation of scarce healthcare resources. For an economic evaluation of healthcare interventions to inform optimal decisions, it is often necessary to understand the effects and costs of an intervention across the lifetime of a patient. In the absence of primary data to inform this, economic models are required to extrapolate beyond observed data, collate best available evidence from disparate sources and conduct experiments that could not be performed in a real-life setting.

As well as allocating resources to the provision of existing interventions, public monies help conduct medical research into potential new interventions that may deliver future health benefits. Given the opportunity cost of investing in research into new interventions, over the provision of existing interventions, policymakers and funders have shown interest in understanding the economic value, or impact, of publicly funded medical research. Based on logic models developed in the research impact literature, the outputs of economic evaluations can be used in models to assess the return on investment from bodies of medical research.

This thesis presents a critical review alongside a portfolio of seven published works concerned with assessing the value of: (a) healthcare interventions; and (b) funding health research. Chapter 1 presents background to contextualise the works and outline the central themes. Chapter 2 explores the overarching methods and contribution to knowledge and Chapter 3 assesses the impact of the portfolio.

The critical review demonstrates the extensive role the methods developed for health technology assessment can play in research impact assessment and the remaining boundaries and challenges. Self-reflection on the contribution to knowledge and impact of the works, combined with formal bibliometric techniques suggest the work has made significant contribution and had identifiable impact across targeting of future research (by centrality or significant contribution to other research), influencing policy (including clinical guidelines), and potential impact on health outcomes (through implemented interventions).

Table of Contents

List of tables, figures and equations

Abbreviations

AAA	Abdominal aortic aneurysm
BRC	Biomedical Research Centre
CPRD	Clinical Practice Research Datalink
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CEAF	Cost-effectiveness acceptability frontier
CM	Conservative management
CPAP	Continuous positive airway pressure
DES	Discrete event simulation
DSA	Deterministic sensitivity analysis
FWCI	Field weighted citation impact
HES	Hospital Episodes Statistics
HTA	Health technology assessment
ICD-10	International Classification of Diseases 10th Revision
ICER	Incremental cost-effectiveness ratio
INMB	Incremental net monetary benefit
IRR	Internal rate of return
KCE	The Belgian Health Care Knowledge Centre
LMIC	Low- and middle-income countries
MAD	Oral mandibular device
MASS	Multicentre Aneurysm Screening Study
NAAASP	National Abdominal Aortic Aneurysm Screening Programme
NHS	UK National Health Service
NIC	Net ingredient cost
NIHR	National Institute for Health Research
NINDS	National Institute of Neurological Disorders and Stroke
NMB	Net monetary benefit
NPV	Net present value
NVD	National Vascular Database
ODI	Overseas Development Institute
ONS	Office for National Statistics
OSAHS	Obstructive sleep apnoea hypopnoea syndrome
PSA	Probabilistic sensitivity analysis
ROI	Return on investment
QALY	Quality adjusted life years
UK	Great Britain and Northern Ireland
US	United States of America
WHO	World Health Organisation

Authors declaration

I declare that this submission is not substantially the same as any previous submission I have made or that I am currently making, either in published or unpublished form, for an award of any university or similar institution. Until the outcome of the current submission is known, the works submitted will not be submitted for any such award at any other university or similar institution.

The work contained in this critical review is mine alone, however co-authorship of the published works contained in the portfolio is explicitly acknowledged. Full details of the contribution I made to these works is summarised in the following sections, with a list of works and candidate contribution. Co-author substantiation of this contribution are given in APPENDIX H: Co-author declarations.

List of published works

1. **Glover MJ**, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England. *British Journal of Surgery*. 2014;101(8):976-982. doi:10.1002/BJS.9528

Referred to as *Paper 1*. Full details of permissions and a copy of the paper are available in APPENDIX A: Paper 1.

Contribution of candidate: Led update of model including input parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), performed cost-effectiveness analyses. Contributed to conception of research question. Led preparation and submitted manuscript. Responsible for manuscript revisions, including additional cost-effectiveness analyses. Listed as first and corresponding author.

2. Thompson SG, Brown LC, Sweeting MJ, Bown MJ, Kim LG, **Glover MJ,** Buxton MJ, Powell JT and the RESCAN collaborators. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. *Health Technology Assessment.* 2013;17(41):1-118. doi:10.3310/HTA17410
   Chapter 7 and Chapter 8 of larger monograph

Referred to as *Paper 2*. Full details of permissions and a copy of the paper are available in APPENDIX B: Paper 2.

Contribution of candidate: Led update of model parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), developed framework and performed structural amendments to operationalised model. Conducted cost-effectiveness analyses, drafted Chapter 7 and 8 and contributed to revisions, including additional cost-effectiveness analyses. Contributed to critical revision of other relevant sections of monograph.

3. **Glover MJ,** Jones E, Masconi KL, et al. Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening. *Medical Decision Making*. 2018;38(4):439-451. doi:10.1177/0272989X17753380

Referred to as *Paper 3*. Full details of permissions and a copy of the paper are available in APPENDIX C: Paper 3.

Contribution of candidate: Co-investigator on National Institute for Health Research Health Technology Assessment grant, involved in conception of research question. Contributed health economics input to conceptual model development and led update of costs and outcomes (including identification and acquisition of data). Conceived, led preparation and submitted manuscript and responsible for revisions. Listed as first and corresponding author.

4. Sharples L, **Glover M**, Clutterbuck-James A, Bennett M, Jordan J, Chadwick R, Pittman M, East C, Cameron M, Davies M, Oscroft N. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and *Health Technology Assessment*. 2014;18(67):1-330. doi:10.3310/HTA18670
   Chapter 4 of larger monograph

Referred to as *Paper 4*. Full details of permissions and a copy of the paper are available in APPENDIX D: Paper 4.

Contribution of candidate: Contributed to systematic review of effectiveness, including data extraction and input to statistical analyses performed for health economic model inputs. Led structured review of literature to inform model amendments and updated model parameters and conducted cost-effectiveness analyses. Drafted chapter 4 and made revisions, including performing additional cost-effectiveness analyses. Contributed to critical revision of other sections of monograph and helped supervise within-trial economic analysis presented in Chapter 2.

5. Raftery J, Hanney S, Greenhalgh T, Glover M, Blatch-Jones A. Models and applications for measuring the impact of health research: Update of a systematic review for the health technology assessment programme. *Health Technology Assessment*. 2016;20(76):1-282. doi:10.3310/hta20760
   Chapter 5 of larger monograph

Referred to as *Paper 5*. Full details of permissions and a copy of the paper are available in APPENDIX E: Paper 5.

Contribution of candidate: Contributed to systematic review in Chapter 3 (including data extraction), led systematic review of economic impact assessment literature (project lead on Chapter 5), drafted Chapter 5 and revisions. Contributed to critical revision of other sections of monograph.

6. **Glover M,** Buxton M, Guthrie S, Hanney S, Pollitt A, Grant J. Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. *BMC Medicine*. 2014;12(1):1-21. doi:10.1186/1741-7015-12-99

Referred to as *Paper 6*. Full details of permissions and a copy of the paper are available in APPENDIX F: Paper 6.

Contribution of candidate: Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and revisions, including additional performing modelling. Contributed critical revision to other sections of manuscript. Listed as first author.

7. **Glover M**, Montague E, Pollitt A, et al. Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes. *Health Research Policy and Systems.* 2018;16(1):1-24. doi:10.1186/S12961-017-0276-7

Referred to as *Paper 7*. Full details of permissions and a copy of the paper are available in APPENDIX G: Paper 7.

Contribution of candidate: Co-investigator on Wellcome Trust led grant, involved in conception of research question. Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and performed revisions, including additional modelling. Contributed critical revision to other sections of manuscript. Listed as first author.

Summary of candidate contribution

The candidate contribution to the portfolio of published works is summarised below. Further information on the selection of works is detailed in Chapter 1 and co-author substantiation of this contribution are given in APPENDIX H: Co-author declarations.

Table 1: Summary of published works and contribution of candidate

Paper (short name)		Year of publication	Authorship/contribution				
			Co-I	1st author	2nd author	Corresponding author	Main analyst*
P1	NHS NAAASP	2014		X		X	X
P2	AAA surveillance intervals	2013					X
P3	Lessons from AAA modelling	2018	X	X		X	
P4	TOMADO	2014			X		X
P5	HTA impact review	2016					X
P6	Estimating returns of cancer research	2014		X			X
P7	Estimating returns of MSK research	2018	X	X			X

*Economic/modelling component

Abbreviations: AAA: abdominal aortic aneurysm; CO-I: Co-investigator on grant funding; HTA: Health Technology Assessment; MSK: Musculoskeletal; NAAASP: National Abdominal Aortic Aneurysm Screening Programme; NHS: National Health Service; TOMADO: Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea

Acknowledgements

I'd like to express my gratitude to my supervisors, Professor Subhash Pokhrel and Professor Steve Hanney. Both have provided invaluable input to the completion of this thesis, helping me navigate the process with their intellectual support and good humour.

The portfolio of works included in this thesis are the result of collaborations with other academics, researchers and clinicians. I am thankful to them all for their specialist expertise, experience and support throughout the projects that produced the papers I have submitted in this thesis. I would like to express my appreciation to external collaborators Professor Jonathan Grant, Dr Timothy Quinnell, Professor James Raftery, Professor Linda Sharples, Dr Michael Sweeting and Professor Simon Thompson who all offered their unequivocal support when approached about submitting these works for my PhD by publication.

I owe much to colleagues past and present, especially those at the Health Economics Research Group, Brunel University London, and in particular Professor Martin Buxton, Professor Julia Fox-Rushby and Dr Louise Longworth. I thank Martin for several helpful conversations we had whilst I tried to mould the PhD proposal, portfolio and thesis.

Thank you to Professor Nana Anokye in the Division of Global Public Health for the positive support as I navigated this around my teaching and other commitments and Professor Heather Gage for the enormously welcome encouragement from my other institution, the University of Surrey.

Thanks to Mum, Dad and Mel: for everything. Lastly, this is dedicated to Kirb.

Chapter 1: Introduction

This thesis presents a critical review of a portfolio of published works with a central theme, considering the methods and application of economic modelling to inform healthcare resource allocation. It reflects on four papers presenting adaptations to decision analytic models evaluating the cost-effectiveness of abdominal aortic aneurysm (AAA) screening and oral mandibular advancement devices (MAD) for the treatment of obstructive sleep apnoea-hypopnea syndrome (OSAHS)[1–4]. In addition, the thesis examines the extension of these methods to three papers on economic models assessing the return on investment (ROI) from health research funding. A literature review considering this relatively new field of research is accompanied by two papers applying ROI modelling methods to research funding in cancer and musculoskeletal diseases[5–7].

Chapter 1 provides background to contextualise the works and begin to explore the central themes that define the portfolio. Chapter 2 extends this to examine and justify the overarching methodologies applied across the works and their contribution to knowledge. In Chapter 3, the impact of the portfolio is assessed using bibliometric techniques supplemented with case studies conducted by desk analysis. Finally, Chapter 4 discusses the conclusions that can be drawn from the critical review, reflecting on the limitations of both the portfolio and this review alongside the potential implications of the works for future research.

1.1 Background

The papers presented in this thesis are concerned with assessing the value of: (a) healthcare interventions; and (b) funding health research. Cost-effectiveness ("value-for-money") can be assessed using economic evaluation; an explicit analytical approach to consider the relative costs and benefits of alternatives. Over the last 30 years, economic evaluation has increasingly been used as a tool to inform the allocation of scarce healthcare resources by governments and decision-making bodies, including the National Institute of Health and Care Excellence (NICE) in the UK.

The aim of economic evaluation is to consider opportunity costs to improve efficiency and maximise the benefits from scarce resources[8]. Opportunity costs refer to the forgone health benefits that could be derived from the next best alternative and represent the trade-offs that exist[9]. Several types of economic evaluation have been proposed to aid the allocation of healthcare resources, which differ principally in the breadth of costs and benefits included and the valuation methods adopted. Cost-benefit analysis is the only method rooted in welfare economics ("welfarism"), concerned with evaluating the effects of resource use on

societal wellbeing (the sum of individual utility) and including all relevant costs and benefits regardless of the sector or individual they impact, valued using willingness-to-pay[10]. In 1996, the US Panel on Cost-effectiveness in Health and Medicine offered an early practical "reference case" to standardise methods, allied closely to this theoretical framework[11]. The Panel's recommendations included some theoretical compromise and recognised the use of the quality adjusted life-year (QALY), a health-specific measure of wellbeing. The QALY is a composite measure of length and quality of life, which supports comparisons across diseases and interventions[12].

An alternative theoretical framework has been proposed focussed primarily on the maximisation of health (QALYs) under budget constraints, referred to as *extra welfarism*[13]. Decision-makers must either know their cost-effectiveness threshold (λ), reflecting the opportunity cost of a QALY or be able to operate under conditions of uncertainty with respect to the value of λ[14]. As a result, some decision-makers have seemingly become QALY maximisers, although this aim may be balanced with other aims and societal preferences[15–17]. The extra welfarist framework which underpins the works is considered further in Chapter 2.

There has been a proliferation of literature performing economic evaluations of new treatments, diagnostics, pharmaceuticals, surgical procedures and preventative measures in recent decades[18–21]. In addition, countries (Canada, Australia) began to formalise economic evaluation as part of health technology assessment (HTA) in the 1980s and 1990s[22,23]. HTA is a multidisciplinary approach to assessing new health technologies/interventions (often pharmaceuticals) on a range of properties including safety, clinical effectiveness and often cost-effectiveness. Additional countries (largely higher-income) have more recently formalised HTA[24–28], although the extent to which cost-effectiveness influences decisions or recommendations does differ and notable exceptions exist, including the US[29–31]. HTA has also expanded in low-and-middle income countries but barriers to widespread adoption remain[32–35].

UK HTA was first formalised in a 1997 UK government white paper which created the National Institute of Clinical Excellence (NICE), later the National Institute for Health and Care Excellence. This gave a statutory body instructions to produce clinical guidance with cost-effectiveness as a core principle for the first time[23]. NICE gave the *extra-welfarist* approach and cost per QALY prominence in its guidance[36]. It has specific preferred methods for analyses through the "reference case", and offers detailed guidance on analytical techniques in its Technical Support Document series[37,38]. Although not formally part of drug

pricing and reimbursement, manufacturers are often required to offer discounts to the National Health Service (NHS), before a positive recommendation can be issued[39]. It is also mandatory for the NHS to make a treatment available within three months following positive recommendations by the Technology Assessment programme[40].

Earlier economic evaluations were often concerned with interventions which were already in widespread use, to justify their continued use, expand usage or to optimise delivery for specific populations (e.g. kidney transplantation and dialysis[41] or coronary bypass grafts[42]). In ex-post analyses, data from observed practice may be available to estimate costs and QALYs for available alternatives, with caveats around population and unobservable characteristics. However, the conceptual problem is different if attempting to compare with the status quo ante (before the interventions), or prospectively assess new interventions ex-ante, as in HTA. The counterfactual becomes a key conceptual problem i.e. What are the consequences of using an intervention and what would have happened in the absence of the intervention?

For ex-ante analysis the counterfactual could be inferred from (potentially matched) observational data on current practice and on the new intervention. Alternatively, randomised clinical trials (RCTs) offer a proxy for the counterfactual by prospectively and randomly assigning individuals to different arms to produce unbiased estimates of incremental cost/QALYs[43,44]. There are practical advantages for collecting data alongside clinical outcomes, in addition to the benefits afforded by these designs.

For an economic evaluation of healthcare interventions to inform optimal decisions, it is often necessary to understand the costs and effects of an intervention across the lifetime of a patient. RCTs do not often follow-up participants for long enough for all the events of interest to have occurred (e.g. costs, mortality) with outcomes right-censored as a result. Acute illnesses with short survival times may provide notable exceptions. In the absence of primary data sufficiently robust and complete to capture all opportunity costs, health economic models (decision [analytic] models) are required to extrapolate beyond observed data, collate best available evidence from disparate sources and conduct "experiments" that could not be performed in a real-life setting[45,46]. Decision models are flexible tools used across multiple disciplines including business, economics and operational research to determine optimal decisions under conditions of uncertainty, based on an axiomatic system[47]. That is that if some condition is met (i.e. the estimated net benefit of a new healthcare intervention is higher than an existing alternative) an action should be taken (i.e. the new healthcare intervention should be funded over the existing alternative).

The increased use of modelling methods has also produced a growing literature on appropriate technical approaches to decision modelling in healthcare[48–50]. In selecting an approach, a trade-off is perceived between simplicity in construction and communication and accurately depicting the decision problem, natural disease history, clinical pathway and outcomes, with some conceptual frameworks more commonly adopted (decision trees and Markov models)[51–53]. In addition, assumptions of process and outcomes are required, not least when faced with an unobserved period in a long-term model. The arising structural uncertainties highlight the importance of establishing model validity, to give some confidence for decision making. This may include external comparison with other studies than those used to populate a model[54,55]. Parameter uncertainty will also affect confidence in results, driven by data sampled from a population and/or choice of model inputs. However, optimal decision making should be based on expected outcomes rather than arbitrary levels of certainty[56,57].

Four papers presented in this portfolio of works are concerned with adaptation of decision models intended to inform UK HTA[1–4]. Given that both stand-alone economic evaluations and those produced for HTA processes have expanded, the practice of adapting existing decision models is increasingly common. Adaptations can increase applicability to a certain population, country or setting, update key model parameters or assess different interventions[58–60]. Some models are constructed with future amendment of input parameters as a core feature (e.g. United Kingdom Prospective Diabetes Study model[61]). When adapting these models, it was necessary to reflect on the conceptual approach, determine structural adaptations required, parametrise and characterise uncertainty. Guidance and checklists have emerged to encourage robust methodology suitable for decision making, including characterising uncertainty and tools to assess model quality and validity, but decisions remain at the analyst's discretion[62–66]. In Chapter 2 the methodology used to adapt existing decision models is explored, with particular reflection on appropriate conceptual frameworks.

As well as allocating resources to the provision of existing interventions, public monies help conduct health research into interventions that aim to deliver future health benefits. The position of this funding in the health research ecosystem is complex, and also comprises private funding and innovation. However, given the opportunity cost of investing in research into new interventions, over the provision of existing interventions, researchers, policymakers and funders have shown interest in the economic impact or return on investment (ROI), of publicly funded health research[67–71]. This can inform considerations regarding future levels

of funding for research, or help understand the mechanisms through which impact occurs and potentially how to increase this impact[72–74].

A fledging field has been recognised in the impact assessment literature, concerned with evaluating the outputs of research, which attempts to use economic models (ex-post) to estimate ROI[75]. Logic models are required to hypothesise the chain of causes and effects, regarding the path of health research from "bench to bedside". Here the issue of the counterfactual discussed earlier is paramount, where research has been funded and new interventions implemented but only one reality is observed. Attempts to quantify the ROI have considered observed increases in life expectancy, which at least in part will be a result of advances in technologies, using willingness-to-pay to value these benefits[76,77]. More recently, novel methods described in this thesis have drawn on the economic evaluation literature. Aggregated health benefits quantified in QALYs have been used to estimate the realised benefits of research advances and equate these with a period of research investment.

Attempts have been made to perform ex-ante ROI analyses either at the aggregate level or for specific programmes of research but require strong assumptions about future usage[78,79]. In the decision analysis literature, value of information analyses have been proposed as an ex-ante method for prioritising research. It compares the cost of a "wrong" investment decision with the cost of resolving uncertainty, to consider whether is it worth conducting further clinical research or investing in improving adoption of cost-effective interventions[80,81]. However, when no clinical studies exist, studies that produce estimates are required before research funders can seek to resolve uncertainty. Though a subtle distinction, the ROI models attempt to quantify the value of eventual realised health gains, including from basic research.

1.2 Selection of the published works

The works, published between 2013 and 2018 (research began in 2011) have been chosen because of the overarching methods and their contribution to knowledge. They share both conceptual and technical approaches to modelling and produce similar challenges in robustly estimating the costs and benefits of respective investments.

Paper 1 (P1)[1], Paper 2 (P2)[2] and Paper 3 (P3)[3] are economic evaluations of AAA screening which use adaptations and extension of existing decision models. The works were sequential and adopted increasingly complex conceptual and structural frameworks to address an evolving decision problem. Paper 4 (P4)[4] offers another distinct clinical application (OSAHS), using a state-transition model, to reflect on these overarching

methodologies. Paper 5 (P5)[5], Paper 6 (P6)[6] and Paper 7 (P7)[7] are concerned with the use of ROI modelling for impact assessment.

P2, P4 and P5 were published in *Health Technology Assessment*, the peer-reviewed journal publishing monographs that accompany all studies funded by the National Institute for Health Research (NIHR) HTA Programme. These reports collate all clinical, statistical and economic aspects of the HTA in full detail. The chapters which were written to report the economic modelling methods and results are the focus of this submission.

I was a co-investigator on research grants which supported the work for P3 and P7 and first and/or corresponding author on four of the seven works (P1, P3, P6 and P7). For the others, I conducted the work with either full responsibility for this contribution (P5) or with substantial analytical autonomy under the supervision of a senior colleague (P2 and P4).

Table 1 provides a summary of the candidate contribution to the published works and APPENDIX H: Co-author declarations, provides substantiation.

1.3 Aims

A portfolio of seven published papers with a central theme are submitted together with this critical review. The aims in this critical review are, as set out in the university's guidance, to:

- Outline the theme(s) that gives the work its defining coherence (Ch.1 & 2)
- Justify the overarching approach and methodologies used (Ch.2)
- Show how the work makes a significant and coherent contribution to knowledge (Ch.2 & 3)
- Provide an assessment of the impact of the work contained in the submission (Ch.3)

The aim of the review is to critically reflect on the portfolio to consider the relationship between research impact assessment and modelling in health economic evaluation, to consider the following research question:

What role can the modelling methods used in health technology assessment to inform the allocation of scarce resources play in informing health research impact assessment?

Chapter 2: Methods and contribution of the published works

This chapter will explore the theoretical position in which the works are couched, the conceptual approaches that inform the analyses and the common methods adopted in the portfolio. Subsequently, the contribution to knowledge of the portfolio is considered, reflecting on the state of knowledge prior to publication as well as the additive nature of the understanding produced between the works.

2.1 A framework for healthcare resource allocation

Analyses presented in P1, P2, P3 and P4 were aligned with the methods of economic evaluation of healthcare interventions most commonly adopted in UK HTA processes by NICE[37]. This approach was also consistent with the methods used to assess ROI of health research funding in P6 and P7. These methods of evaluation are grounded in what is commonly referred to as *extra-welfarism*, although authors have argued about the clarity of this terminology[13,82,83].

Culyer's early exposition of extra-welfarism placed emphasis on the production of health itself (rather than wider utility) and the focus on health care budgets in maximising this stock[13]. NICE's "reference case" represents a specific interpretation of this theoretical framework. Consequently, the methods used to assess returns in P6 and P7 use lifetime QALYs (based on EQ-5D utility scores where possible) as the health outcome of choice and value net health gains using the mid-point of the stated NICE cost-effectiveness threshold[37,84]. The ingrained normative aspects of producing and using subjective measures of health-related quality of life (the "Q") to compute QALYs are present in the ROI analyses in P6 and P7 (as well as P1-P4)[12,16]. The use of the cost-effectiveness threshold to value net health gains reflects the opportunity costs of funding medical research over providing existing healthcare provision within budget constraints[84].

HTA is concerned which comparing interventions and making recommendations at a point in time, though analyses may subsequently be revisited or updated, especially identifying potential cost-effectiveness in sub-populations. However, exploring the returns from health research funding is more often relevant to decisions regarding future aggregate funding of medical research, or understanding the mechanisms through which it is achieved.

2.2 A conceptual approach to impact assessment

Frameworks of impact assessment of health research recognise what can be considered *intermediate* outcomes (e.g. advancing knowledge, capacity-building, further research

development, informing policy) that lead to *final* outcomes - the direct benefits (e.g. increased survival, reduce symptoms of disease, cost savings) and wider economic impacts (e.g. commercialisation, attracting outside investment) of conducting research and implementing new healthcare interventions[85–89]. As with economic evaluation of health care interventions these *final* outcomes are considered of most use in resource allocation considerations[90,91]. To assess the efficiency of producing these outcomes, investments and benefits need to be identified, and valued. However, the methods of economic evaluation of healthcare interventions are relatively established compared to those of research impact assessment, including the specific exercise of estimating the ROI by placing a monetary value on the benefits of research funding[5,85,92,93]. The systematic review in P5 built on a review published in 2004 by Buxton and colleagues[70] and showed that only a small number of additional studies quantifying the ROI of health research had been conducted in the intervening years.

Some of the earliest attempts in the economics literature to quantify the benefits of research expenditure using monetary values focussed on the value of a healthier workforce that results from the introduction of new interventions. Mushkin and Landefeld adopted a human capital approach to value gains from US biomedical research in terms of reduced loss of working time (mortality and disability) and earnings, using market values (i.e. salary/wages)[94]. The authors acknowledged the limitations of doing so: that they tend to overstate the value of lost labour when it can be replaced and neglect benefits in older sections of the population, due to focus on the population of working age.

Wider economic benefits to the economy from funding health research could be considered "spillover" effects from the primary aim of directly improving population health. These spillover may also include the complementary private investment that occurs from publicly funded discoveries[93,95,96]. Until recently, there existed limited healthcare specific estimates for this spillover. A review identified estimates of spillover in other fields of between 20% and 67%[96]. A study by Sussex and colleagues in 2016 quantified this specifically for public funding of health research and estimated an annual rate of return of 15-18%[97].

Later studies have focussed on the inherent value to society from health gains, conducting cost-benefit analyses using willingness-to-pay frameworks to value outcomes and apportion observed changes in outcomes (most commonly improvements in life expectancy) to a period of research investment[76,77,98,99]. A series of studies focussed on Australian medical research have characterised benefits using reductions in Disability Adjusted Life Years (DALYs) and

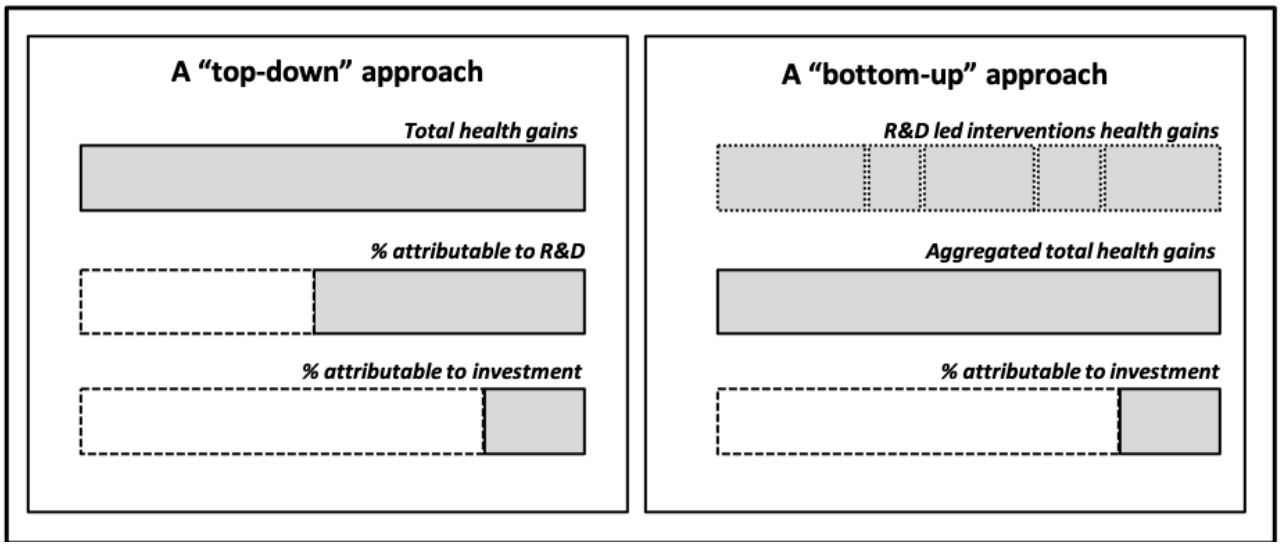
the value of a statistical life year [100–104]. These were more recently extended to focus on returns in spending on the medical research workforce [101]. Some of the earlier studies in this series failed to adequately address the issue of the time lag between investment and outcomes, but later explicitly accounted for this. All these studies adopted a top-down approach: realised (ex-post) or future (ex-ante) aggregate gains were apportioned to a body of research funding.

Johnston and colleagues adopted a different method in a study published in 2006, using a *bottom-up* approach to assess returns from US National Institutes of Health Neurological Disorders and Stroke (NINDS) RCTs completed before 2000[105]. They collated NINDS spending and equated these with hypothetical benefits from implementing new interventions evaluated in the trials over a ten-year period. However, they assumed that the benefit that can be attributed to the NINDS studies was 100%. This is likely to be an overestimate given the need for basic research and other clinical studies to develop new interventions, including from other countries.

A study conducted by colleagues at Brunel University London, RAND and Office of Health Economics, took the *bottom-up* approach further (*Medical Research: What's It Worth?)* [96]. The authors formalised a logic model to relate inputs (UK cardiovascular research spending including basic and clinical research) to outputs (health benefits net the cost of delivery) measured in Net Monetary Benefits (NMB). These models could also incorporate costs savings that may results from the introduction of new interventions. An estimate of the lag between investment and returns and the proportion of the total NMB could attributable to the research investment were required[96]. The ROI analyses presented in P6 and P7 build on this *bottom-up* approach (Figure 1), whereby aggregated health gains from research-led interventions are apportioned to the funding (by estimating the % of the health gain attributable to UK research).

This conceptual approach may offer a more reliable means to quantify returns, with the health benefits pieced together based on the observed introduction of new interventions, based on research evidence. This, at least in part, mitigates some of the attribution issues encountered in other studies, which had often estimated seemingly large returns. However, the approach is by its nature resource intensive (as demonstrated in P6 and P7) and there are naturally a range of simplifications and assumptions (and in some cases compromises) that are required to produce the modelled estimates.

Figure 1: Conceptual approach to attributing health gains (adapted from P5)



The appropriateness of these analytical decisions, as in decision models, will affect the results and their usefulness. Notably, in the modelling in P6 and P7, the scope and boundary of interventions to focus on must be specified. Prioritisation could be criticised as it will lead to missing some interventions that may have been developed and produced health benefits. The aim was to encompass at least the main research-led advances into the model that cover an appropriate breadth of the field being studied. Furthermore, as with the subsequent selection of model inputs, where judgments were necessary, a conservative approach was adopted that would be liable to underestimating, rather than overestimating, returns. Transparency was sought, justification of the approach was provided and modelling assumptions were clearly described.

2.3 Overarching methods of the published works

Each of the cost-utility analyses (P1-P4) were based on interventions which had been assessed before and consequently adapted existing decision models. They focussed on optimising the delivery (P2, P3) [2,3] or investigated cost-effectiveness for specific populations or subgroups (P1, P4)[1,4], incorporating the availability of new data (P1-P4)[1,2,4] and in some instances using different modelling approaches better suited to addressing the decision problem (P2, P3)[2,3]. The works attempted to answer the question of cost-effectiveness with additional nuance: given new data, under what circumstances is the intervention cost-effective?

P5 could be considered an exception amongst the portfolio as it details a systematic review performed to investigate how an ROI, monetary approach, might be applied to assessing the

impact of the UK NIHR funding programme. As such no formal economic model was part of that work. However, P5 reflected on the potential difficulties in adopting these methods for one funder's research spending, drawing on the existing literature (including P6) and methods of literature reviewing constitute a major component of several of the other papers (P4, P6, P7). It therefore forms a useful addition in considering how the methods of decision modelling for health economic evaluation may inform the assessment of research impact.

The work presented in P6 and P7 considers methods to estimate the value of health benefits (and potential cost savings) from research funding. Both the ROI studies (P6, P7) adapted and extended the methods applied to cardiovascular research to quantify the return from investment in cancer related and musculoskeletal related research funding[96].

Table 2 summarises the methodological components of each of the papers in the portfolio, to highlight the shared approaches and explore differences. Key aspects of the works, which will in turn be discussed, have been categorised as:

1. Analysis and outcome metrics
2. Cost perspective
3. Modelling conceptual framework, construction and adaptations
4. Methods for model paramterisation
5. Characterising uncertainty and exploring heterogeneity.

2.3.1 Analysis and outcome metrics

The cost-effectiveness (cost-utility) analyses presented in P1-P4 compare the incremental costs and incremental QALYs of an intervention to appropriate alternatives. If an intervention provides additional benefits at less cost (all relevant downstream costs included) then the intervention *dominates* the alternative. Conversely, if the intervention is both costlier and produces less health benefit then it is *dominated*. The primary metric presented in these analyses is the incremental cost-effectiveness ratio (ICER) - the ratio of the difference in costs and QALYs:

Equation 1: Incremental cost-effectiveness ratio (ICER)

$$[1] \quad ICER = \frac{Cost_1 - Cost_0}{QALY_1 - QALY_0}$$

This is widely used, and the metric recommended by NICE[37] for economic evaluation with the numerator (costs) and denominator (health effects) a well-established convention[11]. The ICER of a healthcare intervention is compared with a decision makers cost-effectiveness threshold, reflecting the opportunity cost of an additional unit of benefit (QALY). An intervention would be deemed cost-effective if the ICER is below the threshold and represents a net benefit over the existing alternative:

Equation 2: Cost-effectiveness (CE) decision rule based on ICER

$$[2] \quad \text{CE if } \frac{Cost_1 - Cost_0}{QALY_1 - QALY_0} < \lambda$$

The ICER is intuitive and interpretable in relation to the opportunity cost of a QALY, but the nature of ratio statistics can be problematic when attempting to characterise uncertainty. This arises because of its non-symmetry and when using sampling methods, the possibility of ICERs with positive and negative values presenting entirely different outcomes. As such, the net benefit statistic has been proposed as an alternative [106], which is a rearrangement of equation [2]. Equation [3] presents the incremental net monetary benefit (INMB):

Equation 3: Incremental net monetary benefit (INMB)

$$[3] \quad INMB = \lambda(QALY_1 - QALY_0) - (Cost_1 - Cost_0)$$

If the monetised value of the health gain exceeds the additional costs of producing the health benefit, then the intervention would be deemed cost-effective (as shown in equation [4]). This also avoids any of the potential confusion which has been recognised when characterising uncertainty, either implementing sampling procedures or when regression methods are used to estimate incremental costs and/or benefits (as adopted in P2, P3 and P4).

Equation 4: Cost-effectiveness decision rule based on INMB

$$[4] \quad \text{CE if } \lambda(QALY_1 - QALY_0) - (Cost_1 - Cost_0) > 0$$

There are also instances when comparing more than two alternatives (as in P3 and P4) where computing appropriate ICERs against the next best alternative is required to appropriately handle opportunity costs. The need to perform a "fully incremental analysis" to compare all available options can become less straightforward, with the introduction of

extended dominance[37]. The INMB avoids this issue altogether, with the option with the highest value against some common comparator deemed the most cost-effective option (at a particular chosen value of λ). The main limitation arises due to the potentially unknown value of λ, but analyses can be run across a range of plausible values.

The ICER was deemed sufficient in P1, where two alternatives were compared (AAA screening vs no screening) and characterising uncertainty did not raise any issues of conflating ICERs with different meaning. However, P2-P4 all used the NMB to present results of competing alternatives, owing to the number of competing alternatives (P2, P4) or to appropriately handle uncertainty (P2, P3, P4). NMB was also crucial in the modelling presented in P6 and P7, as the characterisation of health gains arising from health research. However, rather than modelling the intervention against the next best alternative, an estimate of the INMB against the historical comparator(s) or standard of care when the new intervention began to be used was required.

The main metric in the ROI modelling (P6 and P7), to estimate whether health research investment could be considered good value, was the internal rate of return (IRR). This is more often used in financial analysis to compare the profitability of different investments than for considering healthcare resources. For a stream of investments (health research spending) and returns (net health gains valued at the threshold representing opportunity cost) the IRR is the discount rate that would yield a net present value (NPV) of zero (break-even). This can be compared to a decision-makers discount rate; an IRR greater than the discount rate can be considered a good investment. The IRR is shown in equation [5] below:

Equation 5: Internal Rate of Return

$$[5] \quad \sum_{t=1+l}^{p} \frac{A(NMB_t)}{(1+IRR)^t} - \sum_{t=1}^{p} \frac{I_t}{(1+IRR)^t} = 0$$

Where *NMB* is the net monetary benefit in year *t*, *A* is the attribution of health gains (expressed as a percentage), *p* is the period of health gains considered, *I* is investment (research expenditure) in year *t* and *l* is the time lag (in years) between investment and health gains. The period of health gain considered in P6 and P7 was 20 years and is largely arbitrary, lagged to a period of investment 15 years (P6) and 16 years (P7) prior. This lag was determined using bibliometric techniques applied to guidelines. The average age of citations ("knowledge cycle time") was computed on a set of clinical guidelines driving practice in

each field[107,108]. A period of time from awarding of funding to dissemination of results and a period for clinical guidelines to change practice was added to knowledge cycle time.

There has been some criticism of the IRR, especially in relation to comparing competing investment alternatives and consistency of results from the IRR compared to other alternative measures (e.g. benefit-cost ratio)[109]. However, the IRR might be less vulnerable to these criticisms in the context of P6 and P7 given that the aim is not to decide on investment between cancer or musculoskeletal research, but rather to understand the general magnitude of biomedical and health research returns by using different disease areas as case studies and refine methods. Importantly, the IRR allowed the estimates of the direct health returns and spillover effects to be combined, given the predominance of using the IRR to estimate these returns[96,97].

2.3.2 Cost perspective

The analytical perspective (whose resources) was largely consistent across all works – all broadly considered the UK NHS budget, however there were some distinctions. P1-P4 considered cost-effectiveness of interventions from an NHS perspective, in line with guidance from NICE[37]. The methods in P5 did not adopt a perspective *per se*, given it was a systematic review of ROI models in general. However, discussion did more explicitly focus on the application of these methods to a subset of this budget: the NIHR HTA programme, an NHS arm's length funding body.

In P6 and P7, public expenditure on biomedical and health research was treated as an extension of the healthcare budget; resources that could, at least in theory, be re-allocated to provision of existing healthcare interventions if deemed appropriate. Some organisations fund medical research through charitable donation in addition to, or instead, of government central funding. This provides an additional consideration, on how to conceptually handle these public donations. This was treated as a voluntary increase in the healthcare budget and that the investment of these monies is equivalent to funding received through general taxation.

2.3.3 Modelling conceptual framework, construction and adaptations

P1-P4 used decision models to conduct cost-utility analysis and in P1, P2 and P4 was performed using an adaptation of an existing decision model. P1 and P2 used a model published by Kim and colleagues in 2007[110] and P4 used a model published by McDaid and colleagues in 2009[111]. In P1, the decision problem - a clear identification of the question

being addressed in terms of the intervention, comparator and population - was largely unchanged: is screening men for AAA aged 65 cost-effective? However, the analysis was more sharply focused on the cost-effectiveness of the screening as specifically implemented by the NHS AAA Screening Programme (NAASP) in the UK. In addition, the best available evidence had substantially changed since the original modelling was conducted. This model was primarily based on one RCT[112] and several other large randomised trials had been conducted since, publishing data on growth and rupture of AAA[113].

The introduction of the NAAASP in the UK had produced updated data on key estimates (e.g. community prevalence of AAA, uptake of screening invites in the "real world"). In P2, P3 P4, these works reflect that the decision problem was different from those in the initial modelling exercises. P2 used the same updated data as P1, but required substantial structural amendments to address the decision problem: what are the most cost-effective surveillance intervals for AAA in men aged 65 identified by screening? P3 detailed a de novo discrete event simulation (DES) model which was developed to allow the investigation of diverse range of decision problem regarding AAA screening and its optimisation. The paper reflected more broadly on lessons learned from using state transitions models (as adopted in P1, P2 and P4) for modelling healthcare interventions. The model was subsequently used to assess the cost-effectiveness of screening women for AAA[114]. The decision problem addressed by the analysis in P4 was: are MADs a cost-effective treatment for people with mild to moderate OSAH? The model developed by McDaid and colleagues originally assessed the cost-effective of continuous positive airway pressure devices (CPAP), with MADs one of the comparators under consideration. The model adaptation comprised two objectives as outlined in P4: (i) reflect emerging data since the model was built and (ii) focus on the mild/moderate severity patient population[4]. This included reassessing some of the modelled relationships between OSAH and road traffic accidents, cardiovascular disease and stroke. Data was incorporated from a meta-analysis, including data from the RCT presented in other chapters of P4.

The models used in P1, P2 and P4 adopted a state-transition approach, often referred to as Markov models in the health economics literature[53,115]. These models use distinct health states that individuals can inhabit at any particular time, to chart the movement (transition) between these states over a specified time horizon (timespan of the analysis). These states characterise different conditions of health/disease that an individual may experience, with associated costs and utilities (to quality adjust survival). A set of transition probabilities determine the movement between health states per specified cycle (3-monthly in P1 and P2

and annual in P4). It may be more appropriate to consider these state-transition or semi-Markov models, primarily because they relaxed the assumption regarding the memoryless nature of Markov models. The probability of movement between states did, in some instances, change depending on how many cycles had elapsed or how long a state had been occupied. State transition models are recognised as a useful and commonly adopted conceptual approach to decision models in healthcare to estimate incremental costs and effects (QALYs) and can readily characterise uncertainty[53].

In P2, structural amendments were made to the existing cohort state-transition model to allow for different surveillance strategies, compared to the existing surveillance of AAA depending on the size of the aneurysm. In the original model annual rescans were initiated at first screen if AAA was 3.0cm- 4.4cm (small AAA) and 3-monthly rescans initiated if AAA was 4.5cm to 5.4cm (medium AAA)[110]. An individual was referred for elective surgery if the AAA was larger than 5.4cm (large AAA). The model comprised health states for each of the AAA sizes (small, medium large), as well as related patient pathway including elective and emergency surgeries and death. This corresponded to the surveillance protocol adopted in the randomised controlled trial which also provided much of the data to populate the model, including transition between AAA states[112]. As the model only considered one policy vs no screening, it did not explicitly account for time between intervals, instead using an approximation where transitions could occur in the model in every 3-month cycle (even when rescan was one year). This was also the manner in which the analyses presented in P1 were conducted, which focussed on updated input parameters rather than structural amendments.

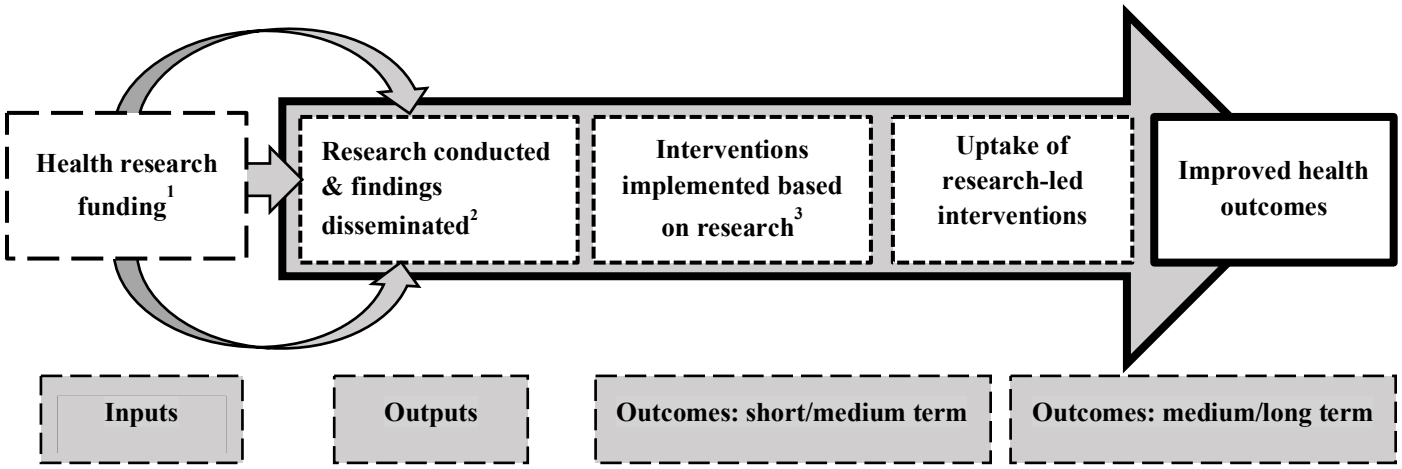
To examine different surveillance intervals involved extensive reprogramming (in MS Excel) of the model through the inclusion of tunnel states. Tunnel states allow further deviation from the Markov assumption. In the model presented in P2, individuals pass through tunnel states (between scans) to determine associated but unobserved transitions between AAA size categories and explicitly allow for the length between rescans. Members of the cohort are "tunnelled" through these other 3-month states, the number of which correspond to the length of intervals (unless transition to death state) before entering back into main health states at their next scan. For instance, for a surveillance strategy where those with AAA 3.0cm to 4.4cm are rescanned every two years, six unobserved tunnel states between scan and rescan are required. This allowed the investigation of different surveillance strategies, but effectively required a new incarnation of the model for each new comparison.

The discrete event simulation (DES) in P3 was developed as a direct consequence of the limitations of the model encountered during the research conducted for P1 and P2. In P3, the

overriding advantages of the DES (built in R programming language) over the "Markov Model" are explored and the case of AAA screening highlights the flexibility DES offers, including the ability to more precisely model the clinical pathway and overcome the constraints of fixed cycles imposed by Markov models, as others have suggested [48,116]. In relation specifically to AAA screening the main three advantages were: (1) Ability to define AAA growth and rupture as a continuous relationship (2) Incorporating heterogeneity in AAA growth rates (3) Ease of evaluating changes to the screening protocol. The latter is the most prominent from the experience of P2, that any number of fixed parameters (e.g. surveillance intervals, threshold for surgical intervention) could be altered and model re-run easily. The notable potential downside is model run time, greatly increased over the Markov Model, given the individual-level simulation adds an additional layer of sampling over the iterations required to characterise uncertainty.

The ROI modelling in P6 and P7 relies broadly on a logic model akin to that presented in the Payback Framework[87] to characterise how health research funding leads to health outcomes. Whilst recognising that the logic models must make simplifying assumptions in conceptualising the chain of cause and effect, the Payback Framework does recognise some of the complexity in how this impact occurs. However, the ROI models do rely on a linear characterisation of the process. Broadly speaking the inputs to the ROI model in P6 and P7 are research funding, the outputs are an evidence base on the effectiveness and cost-effectiveness of an intervention and the outcomes are population health gains (net the cost of delivery), as shown in Figure 2.

Figure 2: ROI models: A simplified process from research funding to outcomes



1.From multiple sources and to multiple researchers or institutions; 2. Both basic and clinical research; 3. Influenced by clinical guidelines and HTA guidance

For the operationalisation of the ROI modelling, this can be considered a continuous process, where new interventions supersede the standard of care, according to available data on usage. The models do inherently make assumptions about the causal nature of research evidence on clinical practice, which has been challenged[117,118]. The models, by their nature, cannot be compared with a counterfactual where the research did not take place. However, in trying to use a *bottom-up* approach, this in part begins to address these issues. In P6, case studies were performed to qualitatively explore the links between evidence and practice[108].

The evolution of pharmacological treatments offers a potential example that illustrates this kind of process. In P7 for instance, in modelling the net health gains resulting from new treatments for inflammatory musculoskeletal conditions (e.g. rheumatoid arthritis) saw the emergence of new generations of pharmacological interventions. At the beginning of the period of health gains modelled (1994) the use of disease-modifying drugs (DMARDS), primarily methotrexate or sulphasalazine, was the standard of care. However, the way this was being used was evolving from part of an escalation protocol towards early and aggressive therapy, informed at least in part by research both in the UK and elsewhere[119–121]. In the late 1990s new biologic DMARDS were licensed in the UK (infliximab being the first) and begun to be used in clinical practice, subsequently also undergoing evaluation by NICE. Over the decade following the launch of infliximab, a number of new pharmaceuticals with similar mechanisms reported favourable Phase III clinical trials - Golimumab was the last included in the ROI model, first used in clinical practice in 2010 (based on usage data).

Not all advances necessarily represent large improvements in outcomes. Health gains from later generations of pharmaceuticals can be less significant compared to improvements observed with early mainstays of treatment. This was recognised in P6 in relation to surgical and radiological intervention and subsequent pharmacological interventions cancer. This may be a consequence of pricing of modern pharmaceuticals, that in some instances is done so to produce estimates of cost-effectiveness close to or at perceived cost-effectiveness threshold of NICE and other similar HTA agencies. This also highlights a further assumption of the models in P6 and P7: that private investment is reflected in the costs borne by the NHS and net against health benefits provided.

In P6 and P7 a de novo model was required. These models were spreadsheet based, operationalised in MS Excel (as models in P1, P2 and P4). As expressed in the IRR presented earlier (Equation 5), the ROI model works on the basis of four estimates:

1. Aggregated health gains of a body of research-led interventions.

2. Time lag between research funding and health gains.
3. Attribution of health gain to research funding.
4. Lagged research funding investment.

To estimate improved health outcomes, and in the absence of being able to directly observe the incremental health gains ex-post, components are required to implement the ROI model: (1) the estimated per individual lifetime health gains (characterised by the NMB) from new interventions over historical comparators (2) uptake of new interventions as a 20-year timeline of usage. The conceptual approach to the first of these is generally easier – studies (primarily model based) that have estimated the lifetime costs and QALYs of the research-led interventions of interest need to be identified. The latter requires identifying sources of usage for older and newer interventions included in the model over the 20-year period. Estimates of the lifetime net health gains per person against historical comparators were multiplied by the number of individuals benefitting from the intervention in each year to produce an estimate of aggregate net health gains. These gains were then apportioned to UK research funding as shown in Figure 1. An estimate of the proportion of the benefit produced by UK health research funding was determined through bibliometric techniques applied to clinical guidelines[108].

In the original *What's It Worth?* study, an existing modelling exercise had identified improvements in survival attributable to treatments in coronary artery disease, which could be used as a starting point[122]. No such comparable publication was found for cancer or musculoskeletal interventions, therefore different approaches were sought to identify research-led advances in cancer (1991 to 2010) and musculoskeletal interventions (1994 to 2013) in P6 and P7. Given the breadth of these areas, prioritisation was necessary, whilst ensuring that most of the health gains were accounted for.

In P6, metrics were used to identify clinical areas where improvements had been observed, supplemented by opinions from a range of cancer clinicians and stakeholder groups on developments in the field. Two sources of data were compiled: 1) reductions in cancer incidence (by type of cancer) 2) cancers for which improvements in 5-year and 10-year survival rates had been observed. Triangulating these data with expert input led to a focus on three areas where new treatments had been introduced to practice: colorectal cancer, breast cancer and prostate cancer (which accounted for nearly 75% of improvements in survival between 1900 and 2009). The intervention set also included screening programmes in cervical, breast and colorectal cancer, as well as smoking prevention and cessation, which was deemed a large contributor to falling incidence of cancers.

In P7, the focus was in a disease where the majority of health gains are likely to be morbidity rather than mortality improvements. In contrast to some cancers, incidence of major musculoskeletal conditions is not generally decreasing[123]. Identifying *important* interventions was therefore more difficult and another novel approach was developed. The scope of interventions also required definition, with musculoskeletal conditions less easily definable. After consultation with experts, the World Health Organisation (WHO) International Classification of Diseases (ICD-10) sub-classification was used. With reference to the subcategories, a shortlist of possible interventions was drawn up for: Inflammatory arthritis (M00–M14), Osteoarthritis (M15–M19), Connective tissue disorders (M30–M36), Back pain and dorsopathies (M40–54) Osteoporosis (M80–82). Studies identifying the burden of MSD disease in the United Kingdom, relevant NICE Pathways, and NICE and National Collaborative Centre (NCC) Guidelines helped to understand the likely largest patient population (and therefore scope for health gains) and identify developments in the field and draw up the list. An iterative process of expert input then refined the shortlist to reach a final inclusion of interventions.

2.3.4 Methods for model parameterisation

Models were parameterised using best available data identified and collated using a variety of techniques. In the evidence based-medicine literature, a hierarchy of evidence is often recognised in relation to assessing clinical outcomes (primarily relating to minimising bias) and this is also of some relevance to decision models. RCTs and systematic reviews (and meta-analyses) are often placed towards the top of this hierarchy. However, in modelling it is also recognised that different study designs and sources may provide more appropriate data for some input parameters or their use be necessitated in the absence of *ideal* data[124,125]. A modeller may have to acknowledge the potential effect of imperfect data and determine how to approach conflicting data or the existence of multiple potential sources of data. Some of these issues may be explored in characterising the impact of parameter and structural uncertainty on model results.

In P1-P4, several key parameter updates were informed by systematic reviews and meta-analyses. AAA growth and rupture rates are crucial parameters in determining the likelihood of associated AAA events and were taken from a meta-analysis of 18 international studies of AAA screening, presented in P2 and elswehwere[113]. Similarly in P4, the effectiveness of MADs compared to conservative management (CM) and CPAP devices was informed by a

meta-analysis of 71 studies, presented in earlier chapters of P4 and elsewhere[126]. In the absence of health-related quality of life data, a subjective sleepiness score widely used in OSAHS studies was mapped to EQ5D and SF-6D utility (using TOMADO data) to use results of meta-analyses in the model.

Whilst not a formal model being parametrised, P5 was a systematic review of the literature (database searching, supplemented by hand searching, snowballing and citation tracking), encompassing extensive search terms and the searching of multiple bibliographic databases (see footnote[1]), consistent with this methodology. The synthesis of these data was narrative rather than a formal statistical meta-analysis.

It is potentially not feasible nor desirable for all input parameters to be informed by systematic review. Parameter input selection should be, as far as possible, transparent and justified. Paisley asserts in a 2016 article that there ought to be a minimum level of searching performed for key parameters in a decision model[127]. Others have also recognised the trade-off that exists between rigour and pragmatism[128,129].

In P1-P4 observational data informed many of the parameters, which was often appropriate given the specific decision problems being addressed or type of input. In AAA modelling (P1-P3), the NAAASP yielded contemporary data (and biggest sample size) on prevalence, size distribution of AAA at first screen at actual costs of the screening, including invitations to screening. These data were also most relevant to the UK decision-making. Arguably, with AAA growth and ruptures rates from meta-analysis and other AAA attributes and key screening parameters coming directly from the NAAASP, the most sensitive parameters were accounted for in the update. These inputs were updated for P1/P2 based on most contemporary data available (2011/12) and subsequently for P3 (2014/15). The detailed paramterisation of the model presented in P3 was published in a larger monograph, concerned more with the application of the decision model than the conceptual advantages of DES[130].

The remaining parameters of note were AAA mortality (including elective, repair emergency repair and rupture mortality) and surgical repair costs. A potential mortality risk is created by the possibility of elective repair and the difference in mortality compared to rupture/emergency repair, as well as associated costs of these events will influence cost-

[1] Ovid MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, Cumulative Index to Nursing and Allied Health Literature, The Cochrane Library, including the Cochrane Methodology Register, HTA Database, the NHS Economic Evaluation Database and Health Management Information Consortium

effectiveness. The definitive UK RCT had largely informed the original model, including a micro-costing of surgical repair[112]. These other key data were sought from UK observational and registry datasets (no formal searching of the literature was performed) in P1/P2 and a separate, though similar, exercise performed for P3. Lack of formal search strategy could be criticised, but the research team included vascular specialists familiar with the field and the issue of contemporaneity was considered of critical importance.

On this basis, data from the National Vascular Database (NVD) and Hospital Episode Statistics (HES) was sought to give most up-to-date data. The NVD was preferred because of the greater granularity in measurement of key surgical cost components including: length of theatre time, and hospital stay (delineated between vascular ward and critical care). Costs were updated by removing components of the overall surgical repair costs and replacing key components with updated mean resource use (as well as cotemporary unit costs). This novel costing practice broke the individual patient link with the data, including estimates of uncertainty around the mean costs, as taken from sampled data. This aspect was not considered ideal, but was deemed preferable from using outdated costs, that would not be reflective of the opportunity costs in practice (e.g. hospital length of stay had fallen). This method also allowed costs to be estimated separately for men and women in subsequent modelling exercise performed using the DES from P3[130].

Whether identified through structured review or more pragmatic searching this relevance (or suitability) was key in parametrising all models. Further structured reviews, which fall short of full systematic review were conducted to identify and update other parameters in P4. These reviews used one bibliographic database (MEDLINE) to try and identify appropriate literature to inform model update. A criterion was pre-specified under which evidence was prioritised:

- evidence was specific to a mild to moderate OSAH population
- estimates were UK specific or more relevant to the NHS
- data were more robust (based on characteristics such as sample size and study design)
- evidence was contemporary compared with previous estimates or
- new evidence facilitated improved modelling (for instance longer-term data or enabling structural improvements) of OSAH and its treatment [4]

In P6 and P7 a similar approach was taken to identify cost-effectiveness analyses that quantified the net health gains of research-led interventions. In P6 the NHS Economic Evaluation Database and MEDLINE were used to identify appropriate studies (although a formal search strategy was not outlined as in P4). A similar criterion was defined to support these selections and methods in the absence of *ideal* data based on a preference for UK specific estimates and in particular analyses presented in NICE TAs and NIHR HTA. In exceptions where international evidence was used, costs were converted using purchasing power parity exchange rates[6]. Analyses from these sources may be imperfect but are likely to be thorough and have experienced substantial scrutiny, especially if used for decision-making. In P7, further focus was directed at estimates produced by NICE and NIHR HTAs either as part of clinical guidelines or formal technology appraisal was emphasised (c.70% of estimates came from these sources).

In addition to the per individual NMB, a timeline of usage of the different interventions is a key parameter in the ROI models presented in P6 and P7. However, it is also not straightforward to identify sources to compile these timelines. In addition to identifying appropriate sources to estimate these inputs, a number of necessary assumptions were required. The estimates needed to reflect the number of new users or recipients, to ascertain the incremental benefits of new interventions over historical alternatives. For many of the interventions this also required identification of usage in a specific population, or indication, for pharmacological interventions. In the original *What's It Worth?* study, the authors had taken a somewhat aggregate-level approach:

> "We computed the numbers of new users each year by subtracting from the numbers of users in each year the numbers of users in the previous year, also accounting for the numbers of deaths from all causes… For simplicity, we assumed a constant annual mortality rate of 1%..."[96]

In P6 and P7 this approach was built on to extend the *bottom-up* nature of the modelling. Rather than relying on trends at the population level, data was sought that would allow the construction of an estimated timeline of *actual* usage, based on available data. However, this was not always possible and population-based estimates were utilised for some interventions (notably smoking prevention/cessation and screening programmes in P6 and treatment of lower back pain in P7). Different techniques were used depending on interventions.

For surgical interventions, HES data on the number of procedures performed in each year (P6 and P7) was used. HES was also used to estimate improvements in management reflected by shorter hospital stay (P7). In addition, the National Joint Registry was used for some musculoskeletal procedures (P7).

For pharmacological interventions, data from the Prescription Cost Analysis (primary care prescribing) and Hospital Prescriptions Audit index (secondary care prescribing) was acquired (P6 and P7). These gave a total cost per year of drugs to the NHS (Net Ingredient Cost [NIC]). If required, these costs were apportioned to different indications/populations based on incidence. Using estimates of the cost of a course of treatment (acquired directly or estimated based on annual costs combined with average treatment duration), the NIC was computed into the number of new users of a treatment per year. Where possible this data, was acquired for every year of the period in question, However, where small gaps did exist, linear interpolation was used (often in early years post-launch and pre HTA assessment). These estimates also accounted for changes in drug prices over the period, including entry of generics to the market. In the absence of NIC data, which occurred for few older drugs, NICE costing templates which give estimated usage were combined with incidence data.

For radiological interventions (P6) data was acquired from National Clinical Analysis and Specialised Applications Team on number of courses delivered. However, these data were only available for one year (at the mid-point of the time period). It was assumed that the proportion of patients receiving radiotherapy was constant and combined with incidence figures to estimate number of courses per year.

For physiotherapy and psychological therapies (P7) data from Clinical Practice Research Datalink (CPRD) was used to estimate incidence of lower back pain. Assumptions from NICE costing templates and clinical opinion were used to estimate proportion of incident cases receiving interventions.

For screening programmes (P6) population estimates of eligible population were taken from Office for National Statistics data (ONS) to reflect the eligible population in each year. The chosen estimate of NMB incorporated the uptake (%) of screening to avoid overestimation of benefits from the whole of the eligible population attending screening.

This highlights the number breadth and scope of different data sources, the various assumptions necessary and the data intensive nature of the ROI modelling studies. Details of the strategies used to produce these estimates was transparent, although may be open to criticism when compared to rigorous standards applied in HTA.

2.3.5 Characterising uncertainty and exploring heterogeneity

Uncertainty can be characterised as both first-order; unexplained random variability, and second-order; parameter uncertainty as a result of estimated quantities, perhaps due to sample information (e.g. data from an RCT)[64]. First-order uncertainty is unavoidable and can be distinguished from heterogeneity, where variability can at least in part be explained by observables (e.g. patient demographics, underlying disease severity). Understanding heterogeneity may help determine populations for which a new intervention is likely to be cost-effective. Heterogeneity was a key consideration of P4 where the cost-effectiveness analysis was concerned with a mild and moderate OSAHS population. The model presented in P3 allows for easy exploration of effects of heterogeneity, through changing fixed parameters (e.g. age at first screen, gender). Second-order uncertainty is largely the focus of methods for quantifying uncertainty in decision models, through use of deterministic and probabilistic sensitivity analysis[131,132], in addition to exploration of structural model uncertainty[133,134].

Uncertainty in results can be explicitly incorporated into the decision, however for optimal decision making the expected benefit should be the primary driver of allocation decisions and not uncertainty[56,135]. However, some exceptions have been suggested, including if there are high sunk costs, or if the incentive to perform more research, to resolve uncertainty, would be removed (as may be the case with positive recommendations and private R&D)[56,57].

In P1, P3 and P4, probabilistic methods were used to characterise uncertainty in model outputs. In HTA, probabilistic sensitivity analysis (PSA) is often considered the gold standard for characterising the joint effect of uncertainty in model inputs on modelled outputs[131]. PSA is a stochastic method, which involves assigning a probability distribution to each modelled input to explore the full range of uncertainty around the best (mean) estimate. A number of simulations are then performed to concurrently sample from each of the input parameter probability distributions and produce a further distribution of model results (ICER or INMB).

This uncertainty was visualised on the cost-effectiveness plane. A probability of cost-effectiveness at a given λ can be computed and visualised using the cost-effectiveness acceptability curve (CEAC) and cost-effectiveness acceptability frontier (CEAF). These assist decision makers in understanding the implications of this uncertainty. The CEAC shows the intervention with the highest probability of being cost-effective at each λ, whilst the CEAF shows the intervention (and associated probability of cost-effectiveness) with the

highest expected net benefit [136,137]. The CEAC was presented in P1, P3 and P4, although the CEAF was only presented in P4. However, this would not have materially altered the interpretation of the results for P1 and P3. The CEAF can often be informative where more than two alternatives exist (as in P4).

In P2, P6 and P7 deterministic methods were used to explore uncertainty of model results, informed by alternative sources of data or uncertainty range in model inputs. Deterministic sensitivity analysis (DSA) involves changing one or more input parameters and observing the impact on results. Key parameters inputs were varied, including the value of λ to reflect historic perceived values and recent empirical estimates in P7[138]. Deterministic methods can be useful to explore the boundaries of uncertainty through scenario analysis (e.g. optimistic, pessimistic) by combining plausible combination of model inputs. The use of "what-if?" analysis may also suggest potential outcomes if parameters values could be influenced e.g. increasing uptake of AAA screening or compliance of MADs would improve cost-effectiveness; decreasing the time lag between funding and health gains would increase the ROI. In P6 and P7, sensitivity analyses suggested that if the time lags were reduced to 10 and 11 years in cancer and musculoskeletal research, the IRR would have been 4.5% and 2.3% higher respectively.

These methods may be used to supplement PSA, or as an alternative where PSA is not feasible. However, its main limitation is that these analyses do not reflect the likelihood of any input parameters value being the true value (as with PSA). Feasibility was the principal reason that PSA was not performed in P2, P6 and P7. The necessary judgments and assumptions in the ROI models, beyond stochastic concerns, combined with the number of studies which did not provide adequate data to inform PSA distributions meant that any uncertainty from PSA would be potentially liable to spurious precision. The application of PSA was considered in particular during the modelling for P7, but as in P6, these issues outweighed the perceived benefits of applying the methods.

Table 2: Overarching methods of the published works

Paper	Analysis and outcome metrics	Cost perspective	Conceptual modelling framework	Model construction/ adaptation	Methods for model paramterisation	Characterising uncertainty
NHS NAASP (P1)	Cost-utility modelling (ICER, NMB)	UK NHS (£)	State-transition decision model (semi-Markov)	Model input parameters	Systematic review Observational data	PSA DSA
AAA surveillance intervals (P2)	Cost-utility modelling (ICER, NMB)	UK NHS (£)	State-transition decision model (semi-Markov)	Model input parameters Structural amendments	Systematic review Observational data	PSA DSA (scenario analyses)
Lessons from AAA modelling (P3)	Cost-utility modelling (ICER, NMB)	UK NHS (£)	Discrete event simulation decision model	De novo model	Observational data	PSA DSA (scenario analyses)
TOMADO (P4)	Cost-utility modelling (ICER, NMB)	UK NHS (£)	State-transition decision model (semi-Markov)	Model input parameters	Systematic literature review and meta-analyses Structured literature review	PSA DSA (scenario analyses)
HTA impact review (P5)	Systematic Review	Health Research Funder (NIHR HTA)	N/A	N/A	N/A	N/A
Returns to cancer research (P6)	Return on investment modelling (IRR, NMB)	UK Public and Government (£)	Logic model	De novo model	Pragmatic literature review Observational data	DSA (scenario analyses)
Returns to MSK research (P7)	Return on investment modelling (IRR, NMB)	UK Public and Government (£)	Logic model	De novo model	Pragmatic literature review Observational data	DSA (scenario analyses)

2.4 Contribution to knowledge of the published works

The contribution to knowledge of the works is summarised in Table 3. It details the state of knowledge before the publication of each of the papers, the limitations of existing literature, research questions unanswered and the contribution of each of the seven works to advancing the field. Some of these contributions have been implicitly considered in previous sections of this chapter. The aim of this section is to more formally explore these contributions and the iterative nature of all of these works, which either directly built on other works in the portfolio or other studies in the field.

P1 and P2 were published as outputs of an NIHR funded project (the RESCAN collaboration) exploring two related questions regarding AAA screening programme for men aged 65: 1) is the NAAASP as implemented cost-effective? 2) can cost-effectiveness be improved through altered surveillance intervals? Previous modelling studies and within-trial cost-effectiveness analyses had shown that a one-off ultrasound screen followed by surveillance of identified AAAs (annually for AAA>3cm, 3 monthly for AAA> 4.4cm) was likely to be cost-effective[110,112,139]. However, data emerging from the NAAASP had shown that prevalence had fallen, those AAA being identified were on average smaller and uptake of the invite to be screened was lower than in RCTs. In addition, previous costs were estimated in 2000/1 prices and new, non-invasive vascular surgery was increasingly being performed (nearly 70% of procedures in 2011/2012)[2]. This raised major concerns over the current cost-effectiveness of the programme and changing surveillance intervals might be one way to improve cost-effectiveness. P1 and P2 adapted an existing model published by Kim and colleagues[110] to address these questions. Analyses presented in P1 and P2 suggest that despite updating key parameters, the NAAASP remained highly cost-effective. Surveillance intervals could be extended for those with smallest aneurysms (1 year to 2 years for AAA 3cm to 4.4cm) at very limited clinical risk, but with associated cost savings from reduced number of surveillance scans.

P3 built directly on the methodological limitations encountered in the modelling presented in P2. In employing a DES conceptual framework to evaluate AAA screening, the paper details the model's ability to easily evaluate further changes to screening protocol and adds reflections to the wider literature considering the use of DES in decision modelling of healthcare interventions. In P3, AAA screening serves as a case study, highlighting the technical advantages of DES over state-transition alternatives.

P4 built on a previous model constructed by McDaid and colleagues to assess the cost-effectiveness of CPAP in OSAH[111]. McDaid acknowledged the limitations of their study regarding the evidence on MADs and made two research recommendations of particular note:

- Further investigation of the effectiveness of CPAP for populations with mild sleepiness is required.
- It remains unclear precisely what type of dental devices may be effective and in which populations with OSAHS. The effectiveness of dental devices compared with CPAP in mild and severe disease populations is unclear[111]

The modelling in P4 directly attempted to address these uncertainties. As described earlier in this chapter, model parameters were updated based on the TOMADO RCT and structured reviews of the literature. The model estimated CPAP was likely to be more cost-effective than MADs in mild/moderate OSAH (although uncertainties remained) and that MADs would be a cost-effective option in those unable to tolerate CPAP.

P5 built on previous reviews of the literature concerned with ROI models in health research impact assessment. It highlighted the limited breadth of the field that exists in modelling the ROI of specific programmes of research funding. The conclusions of this work reflected on the limitations of the small number of studies that had considered programmes of research, relating to assumptions around attribution. These limitations reflect the inherent difficulty in attributing health gains to a specific research programme, especially ex-ante.

In P6, a novel *bottom-up* approach was applied that built upon previous work on ROI modelling of health research funding[96,105]. A de novo model was constructed to bring together key data on NMB and new intervention usage. In the absence of literature to guide the model, detailed methods for prioritizing interventions included in the ROI model were developed. Mortality gains were used as a proxy for overall health gains combined with expert opinion to produce disease sub-categories of focus. New methods to focus on actual usage of new interventions over population trends were explored. The *What's It Worth* modelling study authors highlighted the potential contribution of the work:

> "… different assumptions in our analyses in one or both clinical areas could have produced different (and differing) results… we would need to replicate the approach in a number of areas to see whether the results are more broadly applicable: the default assumption should be that they are not."[96]

In isolation, P6 only partially helps to address these uncertainties around the returns from health research funding. However, when considered together with the ROI work of P7, it suggests a strong common order of magnitude of the historical gains from medical research funding (with IRR of four studies between 7% and 10%). Furthermore, P7 applied detailed *bottom-up* ROI methods to a clinical area where changes in population smoking patterns are likely to have had minimal effect on health gains in the area (unlike CVD and cancer). This directly addressed one of the major research recommendations of P6:

> "It would be valuable to undertake an investigation in another clinical area in which smoking is not important to see whether similar rates of return are found."[6]

2.5 Conclusion

The works share a theoretical position, have adopted similar conceptual approaches and successfully used and adapted economic models to inform HTA and consider impact assessment of returns from health research.

Reflection on the contribution to knowledge of the portfolio of works suggests that the portfolio has substantively built on previous research. Paramterisation methods in P1 -P4 informed estimates of contemporary cost-effectiveness of an implemented screening programme for AAA and the cost-effectiveness of MADs in a mild/moderate OSAHS population. P2 and P3 used additional structural adaptations to reconsider appropriate surveillance intervals and build on these experiences to construct a flexible DES model to assess an array of populations and scenarios.

P5-P7 have shown that ROI methods to health research impact assessment still constitutes a relatively small field, but that it is possible to expand these methods to other clinical areas and taken in combination with other studies, a best estimate of the return on UK health research is around 7-10%.

Table 3: Contribution to knowledge of the portfolio of work

Paper	What was known from previous research?	Limitations, research questions unanswered from previous research in area	Knowledge contribution from portfolio of research
NHS NAAASP (P1)	• Long term modelling suggests screening highly likely to be cost-effective (ICER: c. £3,000)[110] • Markov model validated against external data (from 4-year follow up of MASS)[110] • Subsequent within-trial analysis of 10-year follow-up of MASS RCT also suggests screening cost-effective[139]	• Data from NAAASP suggests both prevalence of AAA and uptake of screening in practice are lower than anticipated in MASS and original model • Is the NAAASP cost-effective based on contemporary data, including use of new surgical techniques for AAA repair?	• Model validated against 10-year follow-up of MASS • Confirms NAAASP likely cost-effective (ICER: c £7,000) • Suggests prevalence can be much lower before programmes appear no longer cost-effective
AAA surveillance intervals (P2)	• A one-off ultrasound scan and subsequent surveillance of AAA either annually (aortic diameter > 3cm) or 3-monthly (aortic diameter > 4.4cm) reduces AAA mortality and is cost-effective (P1) [110,112,139]	• Would alternative surveillance intervals be more cost-effective? • Can the original model be amended to investigate surveillance? • What are implications for risk of rupture and elective surgeries performed?	• Screening intervals could be extended for men with smallest AAA • Most cost-effective intervals at WTP of £20,000 are 2 years for AAA>3cm and 3 months for AAA>4.4cm. • Updated estimates of cost of open and endovascular repair of AAA (2010/11 prices) also used in P1
Lessons from AAA modelling (P3)	• A Markov model was able to evaluate the cost-effectiveness of AAA screening based on contemporary data (P1) • A Markov model could be adapted to explore changes in surveillance intervals (P2)	• Simplifying assumptions were necessary in P2 to model alternative surveillance intervals, including difficulties in conducting PSA • Can the original model (P1 & P2) be successfully re-envisaged as a DES?	• Markov model can be rebuilt as DES and validates against external data • Case of AAA an application to show circumstances under which DES excels (modelling continuous relationships, changing fixed parameters, accurately reflecting timing of events, PSA)

	• DES are likely to be more flexible to model certain clinical characteristics and decision problems but are less commonly used in HTA[48,50] • Screening intervals could be altered to optimise screening programme (P2)	• Are there other aspects of screening protocol that could be changed to improve cost-effectiveness? E.g. AAA size for intervention, including AAA<3cm in programme	• Inclusion of AAA 2.5cm - 2.9cm in surveillance may be cost-effective • Surveillance intervals of 2 years (AAA>3cm), 1 years (>3.9cm) and 3 months (AAA>4.4cm) would be more cost-effective than current protocol • DES can be used to explore any number of changes to the screening protocol, including whether screening women likely to be cost-effective
TOMADO (P4)	• CPAP is highly likely to be cost-effective for treatment of OSAHS[111] • There is limited data on HRQoL in OSAH	• Role of MADs in treatment of OSAHS uncertain[111] • Uncertainty around relationship between OSAH and risk of CVD, stroke and RTAs	• MADs for the treatment of OSAHS are likely to be a cost-effective alternative to conservative management or when CPAP cannot be tolerated (ICER = 6,687 vs CM) • Narrative review suggests "evidence still strongest in supporting role of OSAH in hypertension" • Limited HRQoL data in mild/moderate OSAH exists in literature • Estimates of relationship between ESS and EQ-5D/SF-6D utility • Updated estimates of costs (including MADs) in 2011/12 prices
HTA impact review (P5)	• A limited number of models have estimated returns to health research[67,96,140]	• How has the literature expanded since a review for the NIHR HTA programme was published in 2007?[67]	• Greater clarity of limited breadth of field

	• Some attempts have been made to consider bodies of research funding
	• These studies have predominantly used a "top-down" approach to model health gains and returns
		• Could the NIHR HTA programme (and similar bodies) use ROI modelling to assess impact?	• Nine additional studies considering returns through monetised health gains were identified
			  ○ Three studies considered programmes of research
			  ○ Two used *bottom-up* methodology
			• Attempts to assess impact using ROI models have rarely adequately dealt with issues of attribution
			• The issue of attribution remains the most "significant challenge" with applying ROI methods to NIHR HTA[5]

Returns to cancer research (P6)	• Existing studies have estimated "exceptional" returns with limitations • *Bottom-up* methods can be used to model ROI from heath research[96] • Prior modelling in CVD and mental health had estimated IRR of 9% and 7% respectively[96]	• *What's It Worth* study highlighted that "estimates of the rates of return need to be treated with extreme caution."[96] • Can ROI methods be successfully applied to other therapeutic areas? • Will ROI model produce similar IRR in other research areas? • Challenges exist in identifying research-led interventions and compiling health gains • How best to characterise uncertainty?	• ROI methods able to be applied to another clinical area • NMB of £124bn from new cancer interventions between 1991-2010 • Best estimate of IRR 10% • Smoking prevention/cessation a major contributor to returns • Novel methods developed for prioritising interventions included in ROI model • New methods to focus on actual usage of new interventions over population trends

Returns to MSK research (P7)	• *Bottom-up* methods can be used to model ROI from heath research[96] • Prior modelling in CVD, mental health and cancer had estimated IRR of 9%, 7%[96] and 10% (P6) respectively	• Can ROI methods be successfully applied to other therapeutic areas? • Challenges exist in identifying research-led interventions and compiling health gains • Are returns different in clinical areas concerned with improving morbidity over mortality? • Are returns different in clinical areas where smoking prevention/cessation is not a major contributor of health gains?	• Application of methods in a field where improvements have been in morbidity rather than mortality • NMB in excess of £16bn from new MSK interventions between 1994-2013 • Best estimate of IRR 7% • Novel methods for prioritising interventions included in ROI model • Taken in combination with other studies, IRR around 7-10% for UK medical research • Applying stochastic methods remains a challenge

Chapter 3: Assessment of impact of the published works

In this chapter, the impact of the portfolio of work will be explored using categories from the Payback Framework: knowledge production, benefits to future research, informing policy and health sector benefit[87]. The systematic review presented in P5 identified the Payback Framework as the most frequently used tool in health research impact assessment and it offers a structure from which to consider the impact of the portfolio of work [5]. Bibliometric techniques were used to assess the contribution to knowledge and influence on policy and health outcomes. *Mini* case studies were performed, focussed on a limited number of the academic works identified that cited the portfolio. These examined the influence of the two streams of work, to showcase benefits to future research.

This analysis was supplemented by further qualitative case study "conducted through self-assessment [perhaps] based on desk analysis"[5] to consider broader impacts. Whilst *final* outcomes are of most interest in resource allocation, *intermediate* outcomes are useful in understanding pathways to impact. The final category from Payback Framework i.e. wider economic impacts, is likely to be too difficult to measure and outside the scope of this thesis, but it may be illuminating to consider the challenges of using ROI methods considered in P5, P6 and P7 in relation to P1-P4.

3.2 Bibliometric analysis methodology

Citation analysis offers a means to understanding the contribution to knowledge production and to an extent influence on better targeting of future research, policy and outcomes, through citations in clinical guidelines and policy documents[141,142]. Citation analysis is an attractive tool: it offers simple quantitative measures of potential impact, but should also be interpreted with some caution. A high number of citations does not automatically indicate impact, but may point towards eventual policy influence or health benefits. A high proportion of self-citations may be indicative of targeting of future research, through driving the agenda. However self-citation might be acknowledged as lack of contribution to wider knowledge, or even a more disingenuous attempt to increase citations for the purpose of prestige. The issues of difference in field or journal specific citation counts, citation bias (towards *positive* results) and multi-authored work have also been widely acknowledged[143–145]. Further qualitative analysis can offer insight into use and influence of cited work, although analyst judgement is required.

The citation analysis presented here primarily used Scopus® (correct as of 27th December 2021) to collate data on: number of citations, year of citation and field weighted citation

impact (FWCI). Whilst the journal coverage of Scopus is imperfect (as with other alternatives like Web of Science™) studies have suggested it has wider coverage, whilst also suffering less from issues of accuracy that affect citation counts from search engines like Google Scholar[146,147]. It also contains useful functionality for sorting and interrogating citations and offers additional metrics to assess impact. Weighted citations attempt to offer some context to a citation count, by considering how comparable a number of citations is to other articles in a similar field, or journals[148]. The FWCI considers publications in the first three years after publication, to compute a ratio of total citations to average in the subject field[149]. A value above 1 indicates higher citations relative to other publications in the field. All works in portfolio have been published at least 3 years, so this also gives a comparable metric to compare works published in different years over the period 2013 to 2018. Scopus was supplemented by Google Scholar and Web of Science to identify any additional guidelines, policy documents and grey literature citing the works.

The analysis draws on previous work concerning categorisation of citations and use of bibliometric analysis to move beyond citation counts, towards a deeper understanding of a works broader contribution to knowledge and potential societal impact[142,150]. In three papers (P2, P4, P5) where the economic component was part of a larger study, citations were interrogated to identity what percentage were directly citing the economic analysis. A citation was categorised as being directly related to the economic analysis if it explicitly referred to the methods or results of the economic modelling (or the review of economic modelling of ROI in P5). The number of reviews, discussion or editorials which had cited the work was collated and presented as a percentage of total citations. The number of identifiable clinical guidelines that cited the works were also collated.

To identify the proportion of papers citing a work more than once, total citation occasions in each of the citing papers were counted. This gave an indication of the influence the cited paper was making on the papers referencing it. In addition, a judgment was made on how central or important the work was to the citing article. Methods that have conducted such analyses ordinarily use multiple assessors, and so whilst attempts were made to apply objective criterion, it should be noted that a limitation remains that only one assessor applied the criterion. The guidance offered by Jones and Hanney was adopted in making these judgements[142]. They describe a work as *central* if the key conclusions of the citing article are derived by:

- applying a novel theory, method, scale or technology, etc. set out in the cited article.
- By supporting or developing, either by modification or different application, a concept or method set out in the cited article.
- By refuting a concept or method from the cited article. [142]

To acknowledge works that may not be central to the overall study/article but have notable influence, an additional category was developed. They further categorise a work as making a *significant contribution* if the citing article:

- Describes some aspect of the method, findings or conclusions of the cited article in detail in at least one full sentence and not simply "as Smith has shown" OR "see Smith, 2008".
- Includes a quotation from the cited article [142]

Given the nature of reviews, which can collate broad contributions to a subject, they treated these publications differently to assess importance. The work was considered *important* if:

> "…the cited article is used to help reach or sustain a KEY TAKE-HOME MESSAGE or CONCLUSION of this review/discussion paper…AND that or another citation occasion occurs at a point in the text where a key conclusion or take-home message from the review is being developed or discussed".[142]

The focus of this analysis was on first generation citations, but as part of a case study approach, limited further generations of citations were pursued to try and identify broader impacts. Citations of the seven works are detailed in Table 4 and cumulative citations between are shown in Figure 3.

More detailed bibliometric analysis based on the papers available for full scrutiny is presented in Table 5 and Table 6. Around three-quarters of works identified by Scopus were available for detailed analysis. Two duplicates were identified (P2), one work could not be found based on the reference (P2) and one citation did not appear in the work identified by Scopus once scrutinised (P2). One citation appeared to be an editorial error given the content of that study/report (P4). All other unavailable works were behind a paywall, in journals (or

chapters in books) to which institutional access was unavailable or open access versions were not found.

3.3 Return on investment of the portfolio of works

Given the nature of the publicly and charitably funded research that produced this portfolio of work, the impact could be considered by conducting ROI modelling considered in Chapter 2, at least for the HTA work. The ROI modelling stream (P5-P7) may have had some health impacts, through influence on healthcare research spending, but this would require strong assumptions to quantify the impact, both in compiling health gains and attribution.

Applying the methods to P1-P4, the stream of research funding and identifying time lags could be relatively well accounted for, however aggregating and attributing health gains would be more difficult. The research expenditure which funded the studies that produced P1-P4 is well documented by various funders and academic institutions and is directly traceable to the portfolio. There may be some minor limitations or difficulties related to time spent outside the funded period to complete the manuscript and reports, however this is covered by the University, also predominantly a publicly funded institution. Time lags need not be explicitly modelled as for bodies of research, if the path to health gains can be adequately tracked as a linear process. In theory, if net health gains from implementation or optimisation of AAA screening, or increased use of MADs for treatment could be identified, they could be compiled using the NMB estimated in P1-P4. Although potentially not straightforward, numbers of individuals benefitting from these changes could be identified from public sources to aggregate at the population level (e.g. AAA screening annual reports).

The 2014 Research Excellence Framework in the UK, displayed that it can be difficult to estimate these aggregated health gains. Analysis by Grant, Hinrichs, Gill and colleagues at the Kings College London showed that only 14 submissions were able to compile health gains and 11 monetised these gains[151]. For a sense of an appropriate denominator, 426 REF case study submissions were tagged “health care services” and 325 submissions referred to “clinical guidance”.

When considering the work of one researcher, major obstacle arises when considering attribution, which is no longer being considered at the aggregate. A challenge lies in distinguishing the proportion of health gains that are due to the economic evaluations presented in P1-P4, as opposed to:

- Basic/biomedical research
- Clinical research, including RCTs
- Systematic reviews and meta-analysis of evidence
- Other economic evaluation studies

It may be appropriate in specific circumstances to consider this attribution 100% for some health gains, if they have directly come from modelling i.e. if the AAA programme had changed surveillance intervals because of P3 (there is no evidence this has occurred although it has informed the policy debate). However, this is unlikely to be appropriate. If the consideration was on only my contribution to works that has wider clinical and statistical components, this would add another layer of complexity, although citation analysis might be able to help identify attribution to the economic evaluation component.

3.4 Impact of decision modelling for health technology assessment papers

The most cited paper in the stream of work concerned with decision modelling in economic evaluation was Paper 2 (123), which also had the highest FWCI, suggesting 5 times the expected number of citations (Table 4). However, analysis suggest that 17% of these were directly related to the economic modelling itself (Table 5). Other citations of P2 were mostly related to the meta-analysis of growth and rupture rates of AAA explored in other chapters of the work, which also informed the paramterisation of the economic model. This work was the first of three concerned with AAA screening; P2 was based on a different application of the same conceptual model (cited 73 times) and new modelling in P3 followed from this (and both cited P1).

Of the papers citing these works (P1-P3), over 30% that cited them did so more than once and between 17% (P1) and 33% (P3) were deemed to be central (Table 6). However, given the small sample size for P3 and self-citations (4 of 13 references), this figure is somewhat inflated. However, this may be suggestive of how these works had targeted further research and possibly represents the cutting-edge, as Jones and Hanney explain[142]. These figures also appear relatively high compared to the case studies they performed. P3 has subsequently been used to assess the cost-effectiveness of AAA screening in women[114], revisit optimal surveillance intervals[152] (first considered in P2) and to assess the implications of surveillance affected by the COVID-19 pandemic[153]. The flexibility of this model design has proven to be highly useful is helping address questions regarding provision and optimisation of AAA

screening. Most citations of P3 referred to this finding of the work, often to support a chosen modelling approach, in diverse healthcare applications outside of AAA screening. P1 and P2 have also helped inform discussion on AAA screening policy in the UK as well as set international research priorities[154–156]. In the UK this work has contributed to wider public health debate, including citation in the Chief Medical Officer's 2015 Annual report discussing the health of the "baby boomer" generation[157].

P1 and P2 were central to similar analyses performed in New Zealand[158], Iran[159] and Sweden[160] on assessing the cost-effectiveness of AAA screening for local populations. P1 also contributed significantly to cost-effectiveness modelling of screening in Estonia[161]. Beyond AAA screening, Rossi and colleagues draw on these works (including using compiled cost data) in a recent application to renal ultrasound for kidney cancer screening[162]. Whilst the implications for policy in New Zealand, Iran and Estonia are unclear, and potentially unrealised, the impact in Sweden is more immediately traceable. P1 was cited on six occasions in a literature review performed by the Swedish Agency for Health Technology Assessment of Social Services (Statens beredning för medicinsk och social utvärdering [SBU]). The quality and relevance to a Swedish population was summarised as:

- High quality
- Moderate transferability to Sweden[163]

It was also cited to support the conclusions regarding the impact of prevalence on cost-effectiveness and its use in the discussion suggests the modeling presented in P1, amongst others, had influenced their conclusions (translated from Swedish using Google Translate):

> "The literature search showed that lower prevalence and a higher degree of detection of abdominal aortic aneurysm without screening lead to a significant increase in the cost per QALY at screening."[163]

In the Discussion of the report, the authors make the final statement suggesting that the modeling presented in P1, amongst others, had influenced their conclusions:

> "Conclusion: International studies and model analysis with contemporary Swedish data shows that abdominal aortic screening among men 65 years is cost-effective, with a cost per QALY which is clearly below SEK 500,000."[163]

Based on the systematic review and cost-effectiveness analyses, the National Board of Health and Welfare (Socialstyrelsen) concluded that AAA screening, which has previously been independently and disparately organised at county level, should be offered nationally to all men aged 65[164]. Analysis by Wanhainen and colleagues estimated that the national program would lead to 577 QALYs gained annually[164]. In assessing the impact of my work in relation to these health gains, and considering any attempt to perform analyses of the kind presented in P6 and P7, the issue of attribution is apparent: how much of the health benefit should be attributed to the UK cost-effectiveness analysis presented in P1 and P2, or the numerous international randomised clinical trials that evaluated AAA screening? Or the other international cost-effectiveness analyses cited in the SBU report?

The authors categorisation of study quality and relevance might act as a proxy for extent of influence but this would be hard to quantify in relation to other evidence contributing to the decision (of 10 studies, one other was deemed high quality and moderate transferability, two other high quality and high transferability). The modelling performed based on the local population will no doubt have provided the most compelling evidence. The more granular the unit of analysis (i.e. from body of research [by discipline or country] to specific funder through to an individual researcher[s]), the less appropriate techniques designed to work at aggregate level (like citation analysis of guidelines) become.

P1 and P2 have also been cited in guidelines and position statements produced by professional medical societies in Canada[165,166], Germany[167], Spain[168], Slovenia[169], Pan-European[170] and the United States of America[171,172]. In a 2020 guideline on AAA diagnosis and management, NICE extensively referred to the analysis presented in P2 to conclude that surveillance intervals for the smallest AAAs could be extended, however the guidance deferred to NAAASP protocols (which currently remain unchanged)[173].

Of the four works concerned with economic evaluation of healthcare interventions, P4 was proportionately most cited in review or discussion articles (41%). This may partly reflect the reporting of RCT effectiveness data on MADs subsequently used in systematic reviews and meta analyses. However, 50% of citing articles did directly reference the economic component of the work. In 2021 NICE published a clinical guideline on the diagnosis and management of OSAH. The analyses presented in P4 were summarised in an economic evidence review as “directly applicable” with “minor limitations” and data from the cost-effectiveness analysis helped inform a de novo model constructed for the guideline development[174].

3.4.1 Benefits to future research case study

Whilst the impact of the four works appears diverse, influencing the policy sphere and informing modelling approaches and paramterisation, P3 might be considered most indicative of a collective contribution to knowledge regarding the assessment of health interventions. As discussed, the work advances conceptual frameworks beyond those adopted in P1, P2 and P4, reflecting on the limitations of state transition models adopted in the other works, and potential advantages of DES.

Published articles used P3 to explain or justify the modelling approach or particular aspects of the approach. Weng *et al* did so in an application of DES for improving nursing flow and efficiency of emergency departments[175]. Marrero and colleagues referenced the flexibility, as demonstrated in P3, when choosing a framework to assess treatment plans and genetic tests in cardiovascular disease[176]. Articles detailing original research, discussion papers and books appeared to be influenced when considering the current state-of-the-art and where these methods fit alongside more complex modelling solutions for addressing healthcare assessment and optimisation problems[177–179].

Two articles on the same subject by Tamburis and Esposito[180,181] serve to highlight the incremental nature of knowledge production and scope of influence of the work. The authors extend process mining techniques to DES and use P3 to highlight some of the remaining challenges of successfully utilising DES, including simulation run time and performing adequate validation.

3.5 Impact of return on investment from health research papers

Of those works concerned with return on investment modelling of health research funding, P6 was both the most highly cited (43) and had highest FWCI (4.87) (Table 4). Citations of P6 and P7 related directly to the economic modelling or model results in 60% and 86% of citing instances, respectively (Table 5). On the occasions where other aspects of the work were cited, papers most often referred to the time lags element of the studies and, more rarely, the issue of attribution. These works have made contributions to several subsequent studies and were used to discuss the results of similar exercises in other areas of health [97,182,183] or to inform reviews of methods or empirical estimates of ROI [184,185].

P5 was mostly cited in review or discussion papers (71%); the most of any of the works in the portfolio and may be suggestive of the influence of this work (Table 5). P5 was also used as the basis for two further papers by the same authors, focussed on an accessible summary and a synthesis of studies concerned with multi-project programmes of research[85,186]. These

works have subsequently been cited 121 and 22 times respectively. Furthermore, papers citing P5 often did so more than once (65%) (Table 6). Four papers cited P5 seven or more times, with the most being 27 in a paper that drew heavily on the findings to consider "Research Impact made by Universities of Applied Sciences"[187]. Millar et al used the literature review to contextualise their study considering the assessment of HTA organisations:

> "Another useful resource is Raftery et al. which gives a detailed description of the various methodologies which have been deployed to measure the impact of health research with a view to evaluating the impact of HTA."[188]

Beyond academic citations, P5 has been cited in a policy report published the WHO exploring the institutionalisation of HTA[189] and a policy briefing of the Overseas Development Institute (ODI), concerning evidence-informed-decision making in LMICs[190], suggesting the work may be helping to shape evaluation processes further afield than the UK.

P6 and P7 received coverage in blogs and news articles in the medical research field including in the British Medical Journal[191–193]. These works have extended reach into the sphere of policy debate. The findings from P6 were showcased at an event in June 2014 of the All-Party Parliamentary Group on Medical Research in London, UK ("A Healthy Future for UK Medical Research") alongside a policy briefing document[194,195]. Following the publication of P7, the findings from the stream of work that started with the 2008 publication[96] were synthesised into another policy briefing document for wider circulation[196] and the Medical Research Council (part of the consortium of funders for P6 and P7) highlighted P6 and P7 in its annual report 2017/18[197].

The Belgian Health Care Knowledge Centre (KCE), a semi-governmental institution conducting HTA, included a detailed summary of the findings in a report considering the value of practice-oriented clinical trials[198]. P6 and P7 were referenced in a Cancer Research UK (CRUK) submission to the Business, Energy and Industrial Strategy Committee of the UK Parliament regarding the impact of the coronavirus pandemic on businesses and workers in 2020[199]. P6 was similarly used in a submission by the Alan Turing Institute to the Science and Technology Select Committee in 2017[200]. Both used the findings on the IRR as advocacy for funding of medical research or to substantiate the contribution of UK science to society.

In the UK, the findings of P6 and P7 have been acknowledged at governmental level, although it is less clear how this may have affected government policy. The Secretary of State for Health commented on P6 in 2014:

> "Innovation is essential for improving treatments and finding new cures that can make a difference to patients, and this report is more evidence that investing in UK medical research has wider economic benefits."[201]

The UK Department of Health commented on the findings of P7 in 2018:

> "The findings from this study demonstrate convincingly the value of research funded by government and charities in securing improved quality of life outcomes for patients and in delivering wider economic benefits".[202]

3.5.1 Benefits to future research case study

As highlighted above, all three works appear to have informed understanding of approaches to research impact assessment. In academic publications, more than one of the works were often cited together, suggestive of collective methodological influence of the stream of work on assessing returns from health research. In addition to supporting use of particular methods, some studies used the works to contextualise alternative economic methodology.

Meghea and colleagues (citing P6, P7) conducted an input-output analysis of NIH research centre, contextualising the results and contrasting methods[203]. Similarly, Smith et al (citing P5, P6) performed a macroeconomic assessment of NIHR Biomedical Research Centres (BRC)[182]. They aimed to address some of the concerns of macro approaches highlighted in P5-P7, contrasting their methods with those in P6. They also suggested P6 as a validation of the scale of returns they estimated. Both these studies, primarily focussed on what would be considered spillover effects in P5-P7.

These works have also had a more direct influence on adopted methods. A report produced by researchers at the King's Fund and University of York gave further attention to returns from investment in NIHR BRC[204]. They considered an input-output analysis and net present value model combined with ROI methods aligned with the approach advocated in P6 and P7. The authors revisited the published models to re-estimate the IRR based on different assumptions regarding the value of a QALY. They combined it with other components of

economic value (private and public) to estimate a net present value to UK economic from marginal spend by NIHR BRC.

A collection of studies appears to have been influenced by all three of the works. A series of articles by Hanna and colleagues (including the author's PhD thesis) refer to the work to explore the field and draw on the methods to estimate ROI[184,205]. They adapted the methodology to estimate returns on investment from a single, but major, RCT in colorectal cancer that was thought to have made a large impact on clinical practice:

> "Acknowledging the strengths and limitations of these analyses, an adaptation and combination of the approaches by Glover et al and Brown et al were selected to test the impact of the SCOT trial."

3.5 Conclusion

The portfolio has achieved impact through contribution to knowledge, targeting of future research, influence on policy-making and potential health impacts. The works appear to have made contributions through extensions of methods as well as applied estimates of cost-effectiveness and ROI from health research funding. Impact from P1-P4 was observed in influence on policy-making, though these impacts have not been quantified.

Any *final* outcomes associated with the ROI work are harder to discern, but the acknowledgement of the findings at governmental level suggest potential influence, alongside inclusion in discourse of organisations in the policy-making sphere. Consideration of ROI of the portfolio illustrates the difficulties of applying these methods at a more granular unit of analysis.

Table 4: Citations of portfolio of works identified by Scopus (by year of citation)

Paper	Year of publication	Citations*							FWCI†
		<2017	2017	2018	2019	2020	2021	Total	
NHS NAAASP (P1)	2013	21	10	12	8	7	15	73	5.02
AAA surveillance intervals (P2)	2014	51	15	23	9	9	16	123	5.25
Lessons from AAA modelling (P3)	2018	0	0	2	3	4	4	13	1.69
TOMADO (P4)	2014	8	2	6	5	4	3	28	1.65
HTA impact review (P5)	2016	1	4	7	7	8	9	36	2.05
Returns to cancer research (P6)	2014	15	7	6	6	5	4	43	4.87
Returns to MSK research (P7)	2018	0	0	2	1	4	3	10	1.30

*Correct as of 27th December 2021; †Field Weighted Citation Impact

Table 5: Bibliometric analysis of portfolio of works – type of citing document and use of economic analysis

Paper	Citing articles available for analysis (% of total citations)	Paper specifically cited the economic analysis (%)	Review/ discussion/ commentary (%)	Cited in guidelines	
				Number	Countries
NHS NAAASP (P1)	58 (79%)	NA	24%	6	5
AAA surveillance intervals (P2)	93 (76%)	17%	10%	7	6
Lessons from AAA modelling (P3)	12 (92%)	NA	0%	0	0
TOMADO (P4)	22 (79%)	50%	41%	1	1
HTA impact review (P5)	34 (94%)	21%	71%	NA	NA
Returns to cancer research (P6)	37 (86%)	86%	19%	NA	NA
Returns to MSK research (P7)	10 (100%)	60%	40%	NA	NA

Table 6: Bibliometric analysis of portfolio of works – citation occasions and importance of work being cited

Paper	**Citing papers available for analysis (% of total citations)**	**Papers with > 1 citation occasion**	**Papers where work made a *significant contribution***	**Papers where work was *central/ important***
NHS NAAASP (P1)	58 (79%)	45%	16%	17%
AAA surveillance intervals (P2)	93 (76%)	40%	17%	17%
Lessons from AAA modelling (P3)	12 (92%)	33%	0%	33%
TOMADO (P4)	22 (79%)	41%	27%	9%
HTA impact review (P5)	34 (94%)	65%	24%	21%
Returns to cancer research (P6)	37 (86%)	41%	22%	16%
Returns to MSK research (P7)	10 (100%)	20%	50%	10%

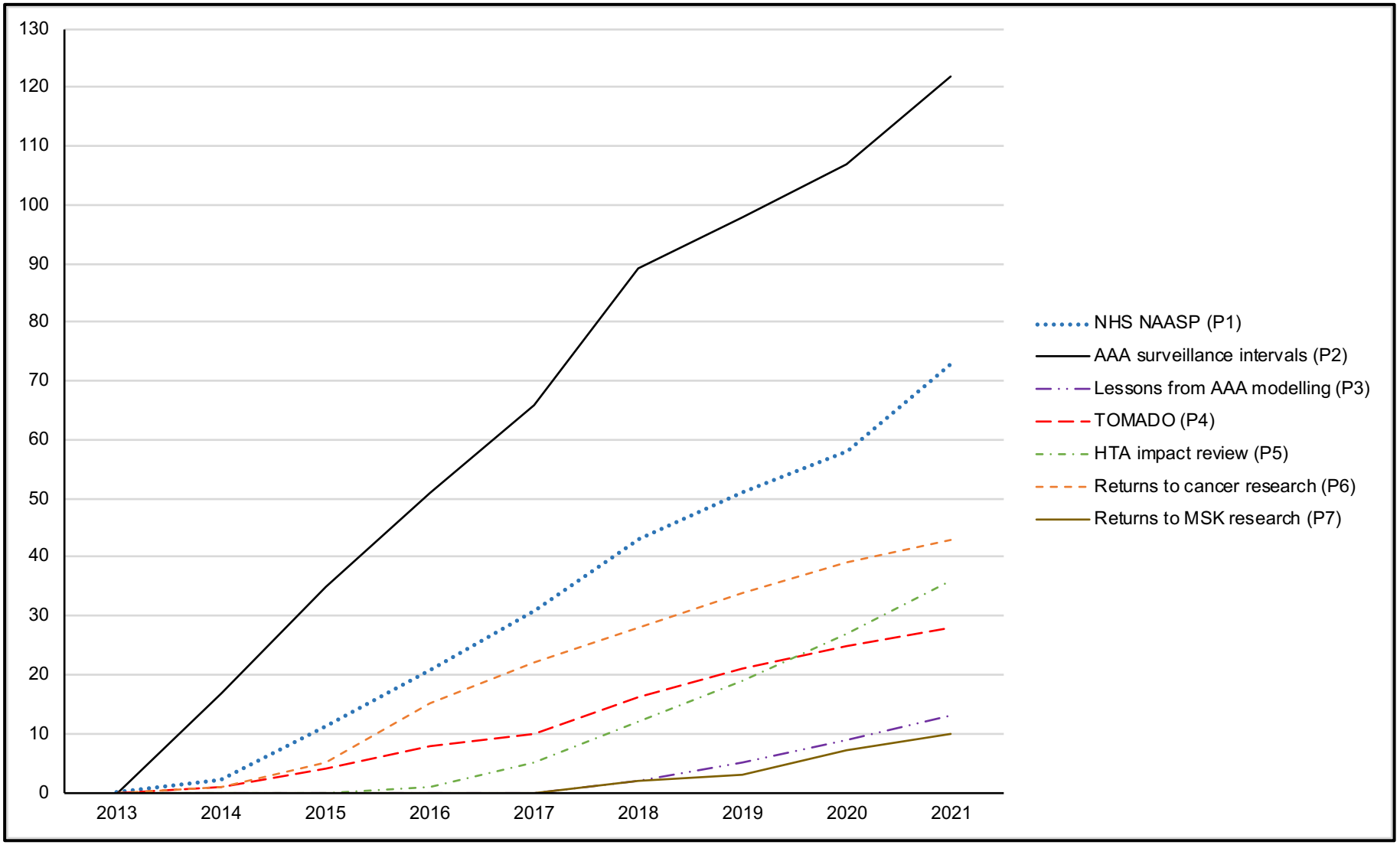


Figure 3: Cumulative citations of portfolio of works 2013-2021 Source: Scopus

Chapter 4: Discussion

4.1 Reflections on the portfolio of works

This thesis has presented a portfolio of seven works from two complementary streams of research with a focus on healthcare resource allocation, which share an underpinning theoretical position and shared methods. The methods of decision modelling highlighted in P1-P4 have shaped the ROI models presented in P6 and P7. Notably, NMB is used as a common measure of health benefit, with lifetime QALYs valued using the cost-effectiveness threshold to reflect the opportunity cost of investing in health research over existing interventions.

In HTA, decision models are most often direct aids to allocation of healthcare resources and concerned with determining the value of adopting a new intervention ex-ante. P1-P4 show the usefulness, and methods, of ex-post adaptations to explore factors which influence cost-effectiveness (population or intervention characteristics) parametrised using observational data, or supplemented by further RCTs and data synthesis. Impact assessment was conducted ex-post, to assess bodies of funding which had produced research-led interventions known to have been implemented. This also appreciates that the process of modelling the ROI from health research also facilitates understanding of mechanisms of realised impact rather than only the impact itself[85,88].

All analyses focused on resource allocation in the UK, taking an extra-welfarist approach and drawing on methods widely adopted and advocated by bodies such as NICE. This raises issues of transferability to other decision-making contexts. Other aspects of *value* could in theory be incorporated into model payoffs (i.e. costs, QALYs), consistent with resource allocation decision making in another jurisdiction. Transferability of results from economic evaluations has been given attention in the literature. This has included reflections on the usefulness of decision models to facilitate this, explanations for variability in cost-effectiveness estimates and which model inputs need attention to satisfy minimum requirements of adaptation [206–209]. To an extent, the model development demonstrated in P1-P4 deals with analogous issues. In the models presented for AAA and OSAH, key parameters such as prevalence or uptake of screening in AAA or compliance in MADs can amended, which is highly likely to be heterogenous across populations. The direct use of the model presented in P1 to conduct analysis in a New Zealand population [158] and models in other settings, which referenced model structure and used model inputs, further demonstrates this[159,160].

However, whilst methods adopted in P1, P2 and P4 show that changing key inputs in state-transition models is conceptually relatively simple and limited structural adaptations are possible, the need to reflect local clinical pathways, procedures (e.g. screening protocols, surgical AAA size thresholds) and adequately handle uncertainty give further support to the advantages of the DES presented in P3. The usefulness of the DES has been further demonstrated to explore other complex decision problems [114,152,153]. In the interests of parsimonious model development, these advantages are unlikely to hold across equally all disease areas and whilst model structural considerations should not necessarily be made on the basis of data availability, the DES does require a richness of data to fully exploit its advantages.

Results presented in P6 and P7 are unlikely to be transferable to other countries, given the complexities in research funding, population costs and benefits, time lags and utilisation of research led advances. This has been previously acknowledged across disease areas, although the results from P6 and P7 alongside previous estimates demonstrate a similar order of magnitude for UK research[96,210]. Despite this, the methods are applicable in other contexts, even if other measures of benefits (beyond or instead of QALYs) are sought for inclusion. However, challenges may arise where HTA methods are less formalised, as there is likely to be more heterogeneity of analytical perspectives and methods applied in the literature than those conducted in a UK context. In this sense, building ROI models which have a footing in a consistent theoretical position may be more difficult. Where new interventions are not routinely subject to any form of HTA, a more fundamental limitation may also present in the availability of data from studies that have adequately estimated lifetime costs and effects for inclusion in ROI models.

Specific differences between the two streams relating to paramterisation, uncertainty and the metrics used to quantify value are apparent. It could be argued that the methods applied to identify data for ROI models fall short of requirements expected of HTA. However, impact assessment is a time-consuming and resource intensive process - each of the studies took approximately 18-24 months to complete. These issues were highlighted in the review presented in P5, and have been recognised in detailed case studies of impact or time lags, over more simplified linear characterisation of processes[117,210,211]. There have been difficulties in incorporating probabilistic sensitivity analyses to consider uncertainty, primarily due to lack of data in some instances as well as the necessary assumptions which break the direct links in uncertainty estimates, as was sometimes necessary in estimating NMB against historical comparators. A degree of pragmatism is necessary in producing these

estimates and any attempts to provide policymakers with more of these data will likely recognise this issue, whilst containing appropriate caveats (as provided in P6 and P7).

However, these methods could add significant understanding about uncertainty in ROI model outputs. In theory, methods akin to those employed in HTA, which sample from probability distributions could be applied to four key ROI model input parameters (research funding, time lags, attribution, NMB). Funding might be considered deterministic, notwithstanding data gaps, as these are closer to a 'known' parameter. Time lags and attribution are estimated mean values, with measures of precision that could be used to inform sampling. The stream of NMB remains the most difficult proposition, where studies often do not give adequate information to define sampling distribution properties. Furthermore, issues remain around the ability to sample NMB directly, or instances where disaggregated cost and QALY data are used to construct estimates of incremental net benefit. Methods would ideally need to account for correlation and potentially employ multiple conditional probability distributions in some parts of the NMB modeling.

Reflection on the contribution to knowledge demonstrates that the works have contributed to applied questions around cost-effectiveness, as well as methods for both adapting and conceptualising decision models in HTA and ROI modelling for impact assessment. The review in P5, and further consultation of the literature, including citation analysis, shows the ROI modelling remains a small field, although interest in the methods continues to grow. Examination of using ROI methods considered P5, P6 and P7 to P1-P4 suggests those methods are more suited to aggregate level over individual investigator analyses. It is likely that logic model simplifications and assumptions required to conceptualise cause and effect work better *on average* at scale (e.g. a disease area) but less easily at more granular level (funders, programmes or individual researchers).

The assessment of the portfolio's impact using the Payback Framework[87], demonstrates disparate impact across knowledge production, targeting future research, influence on policy (with P1, P2 and P4 cited in clinical guidelines) and potential impact on health outcomes. However, the latter component is harder to evaluate and attribute to an individual researchers' body of work. Novel categorisation of citations building on work by Jones and Hanney[142] in particular, was able to show impact beyond simple citation counts. In advancing knowledge and driving future research, each work builds on previous modelling exercises and/or other works in the portfolio, to contribute to methods and application of decision analytic models and ROI models. The model redesigned as a DES in P3 allows flexible

modelling of AAA screening and the advantage of this approach have already yielded further research[130,152,153].

4.2 Strengths and limitations

The preparation of this thesis has allowed in-depth consideration of the two streams of work, to explicitly consider the relationship between modelling in HTA and impact assessment. Whilst other studies may have reflected on some of these issues, the experience of conducting these works across a similar time-frame and attempting to draw on the shared methods prospectively, before then retrospectively examining these challenges suggests this is likely to be a unique study. However, it was obviously not the intention to do so at the beginning and the synthesis may have been approached differently if it were the case.

P1-P4 applied rigorous techniques to adapt existing models to synthesise evidence, populate decision models and make structural amendments, even re-envisaging existing models using an alternative conceptual model. This included structured reviews and extensive use of appropriate observational data to provide contemporary results for robust decision-making as well as considering factors influencing cost-effectiveness.

If data is available, decision analytic models can be validated against external data (as in P1, P2, P3) but assessing the validity of ROI models could be considered more difficult, given the issue of the counterfactual in ex-post analysis. However, P6 and P7 were to some extent exercises in assessing validity, by using the *bottom-up* ROI methods first applied in CVD and mental health research funding to cancer and musculoskeletal research. Each of these new disease areas required the ROI model to be approached differently and the application of novel methods to prioritising interventions, and parameterisation. The results of these ROI models imply similar returns across disease areas.

Whilst assessing impact, the analysis was hindered by a high proportion (6-36%) of the literature being unavailable largely due to pay-walls on academic journals or citations in books (Table 5 and Table 6). Texts were largely confined to those in the English language, although limited clinical guidelines and policy document in other languages were explored. The categorisation and hand-counting of citations was resource intensive and potentially influenced by the single reviewer. This self-assessed process highlights the potential utility of new tools such as Scite, which use deep learning models to specifically consider citation context[212,213].

4.3 Implications and future research

The writing of this thesis has coincided with potential changes in HTA and broader scientific discovery. NICE recently published an updated methods guide, which includes severity weighting of QALYs[214]. This will have implications for health maximisation and models may need to characterise the arising uncertainty. It may also lead to ex-post reassessment of previous HTA; the implications on results of P1-P4 for example are unclear. Over the last 10 years, HTA in the European Union (EU) has been increasingly subject to cooperation through the European Network of Health Technology Assessment[215]. This has culminated in recent regulation for joint assessment – one HTA for all EU countries[216]. This process will potentially still need to account for important country-specific parameters, and historically diverse underlying frameworks for HTA with potential implications for flexibility and model selection.

The issue of time lags is important in ROI models however there may be routes to speed-up basic science to clinical application[217]. Although an emergency/crisis situation, time-lags have been reconsidered in relation to development of COVID-19 vaccines[218]. If lessons from this process can be successfully leveraged, it has possible implications for return from future health research.

The ROI methods presented in P6 and P7 are only able to address historical average returns and do not help to understand how marginal changes in research funding impact outcomes. Future research should give more attention to whether these models can be adapted to consider the margin. They focus on ex-post assessment of the magnitude of returns and understanding impact. Some previous attempts to assess ex-ante have been conducted more focussed on prioritisation of specific research, though have received limited attention[78,219]. Reconsideration should be given to the use ROI models might have in ex-ante assessment and more explicit prioritisation.

The use of stochastic methods for characterising uncertainty should be given more attention in ROI models, potentially with smaller sub samples of data requirements, to investigate feasibility and ways to handle issues such as correlation.

4.4 Conclusion

The analysis conducted in this critical review has shown the extensive role that the modelling methods used in HTA can play in informing health research impact assessment, both conceptually and by sharing technical approaches. This work may be relevant to (a) multi-disciplinary researchers interested in these overarching methods; (b) healthcare decision-

makers and wider stakeholders who consider the application of these methods; and (c) funders involved in commissioning research to reflect on how best to demonstrate impact.

References

1. Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England. *British Journal of Surgery*. 2014;101(8):976-982. doi:10.1002/BJS.9528
2. Thompson SG, Brown LC, Sweeting MJ, et al. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. *Health Technology Assessment*. 2013;17(41):1-118. doi:10.3310/HTA17410
3. Glover MJ, Jones E, Masconi KL, et al. Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening. *Medical Decision Making*. 2018;38(4):439-451. doi:10.1177/0272989X17753380
4. Sharples L, Glover M, Clutterbuck-James A, et al. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airwa. *Health Technology Assessment*. 2014;18(67):1-330. doi:10.3310/HTA18670
5. Raftery J, Hanney S, Greenhalgh T, Glover M, Blatch-Jones A. Models and applications for measuring the impact of health research: Update of a systematic review for the health technology assessment programme. *Health Technology Assessment*. 2016;20(76):1-282. doi:10.3310/hta20760
6. Glover M, Buxton M, Guthrie S, Hanney S, Pollitt A, Grant J. Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. *BMC Medicine 2014 12:1*. 2014;12(1):1-21. doi:10.1186/1741-7015-12-99
7. Glover M, Montague E, Pollitt A, et al. Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes. *Health Research Policy and Systems 2018*. 2018;16(1):1-24. doi:10.1186/S12961-017-0276-7
8. Drummond M, Sculpher M, Claxton K, Stoddart G. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford University Press; 2015.
9. Palmer S, Raftery J. Opportunity cost. *BMJ*. 1999;318(7197):1551-1552. doi:10.1136/BMJ.318.7197.1551
10. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. Welfarism vs. extra-welfarism.

Journal of Health Economics. 2008;27(2):325-338. doi:10.1016/J.JHEALECO.2007.07.003

11. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276(15):1253-1258. doi:10.1001/JAMA.1996.03540150055031
12. Weinstein MC, Torrance G, McGuire A. QALYs: The Basics. *Value in Health*. 2009;12(SUPPL. 1):S5-S9. doi:10.1111/J.1524-4733.2009.00515.X
13. Culyer AJ. THE NORMATIVE ECONOMICS OF HEALTH CARE FINANCE AND PROVISION. *Oxford Review of Economic Policy*. 1989;5(1):34-58. doi:10.1093/OXREP/5.1.34
14. Brouwer W, van Baal P, van Exel J, Versteegh M. When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *European Journal of Health Economics*. 2019;20(2):175-180. doi:10.1007/s10198-018-1000-4
15. Schwappach DLB. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expectations*. 2002;5(3):210-222. doi:10.1046/J.1369-6513.2002.00182.X
16. Williams A. QALYs and ethics: A health economist's perspective. *Social Science & Medicine*. 1996;43(12):1795-1804. doi:10.1016/S0277-9536(96)00082-2
17. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics*. 2005;14(2):197-208. doi:10.1002/hec.924
18. Neumann PJ, Thorat T, Shi J, Saret CJ, Cohen JT. The changing face of the cost-utility literature, 1990-2012. *Value in Health*. 2015;18(2):271-277. doi:10.1016/j.jval.2014.12.002
19. Panzer AD, Emerson JG, D'Cruz B, et al. Growth and capacity for cost-effectiveness analysis in Africa. *Health Economics*. 2020;29(8):945-954. doi:10.1002/hec.4029
20. Neumann PJ, Greenberg D, Olchanski N V., Stone PW, Rosen AB. Growth and quality of the cost-utility literature, 1976-2001. *Value in Health*. 2005;8(1):3-9. doi:10.1111/j.1524-4733.2005.04010.x
21. Warner KE, Hutton RC. Cost-Benefit and Cost-Effectiveness Analysis in Health Care: Growth and Composition of the Literature. *Medical Care*. 1980;18(11):1069-1084. doi:10.1097/00005650-198011000-00001
22. Sculpher M, Drummond M, O'Brien B. Effectiveness, efficiency, and NICE. *British Medical Journal*. 2001;322(7292):943-944. doi:10.1136/bmj.322.7292.943

23. Hutton J, Maynard A. A NICE CHALLENGE FOR HEALTH ECONOMICS. *Health Economics*. 2000;9(2):89-93. doi:10.1002/(SICI)1099-1050(200003)9:2
24. Allen N, Walker SR, Liberti L, Salek S. Health Technology Assessment (HTA) Case Studies: Factors Influencing Divergent HTA Reimbursement Recommendations in Australia, Canada, England, and Scotland. *Value in Health*. 2017;20(3):320-328. doi:10.1016/j.jval.2016.10.014
25. Akehurst RL, Abadie E, Renaudin N, Sarkozy F. Variation in Health Technology Assessment and Reimbursement Processes in Europe. *Value in Health*. 2017;20(1):67-76. doi:10.1016/j.jval.2016.08.725
26. García-Mochón L, Espín Balbino J, Olry de Labry Lima A, Caro Martinez A, Martin Ruiz E, Pérez Velasco R. HTA and decision-making processes in Central, Eastern and South Eastern Europe: Results from a survey. *Health Policy*. 2019;123(2):182-190. doi:10.1016/j.healthpol.2017.03.010
27. Angelis A, Lange A, Kanavos P. Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *European Journal of Health Economics*. 2018;19(1):123-152. doi:10.1007/s10198-017-0871-0
28. Teerawattananon Y, Rattanavipapong W, Lin LW, et al. Landscape analysis of health technology assessment (HTA): systems and practices in Asia. *International Journal of Technology Assessment in Health Care*. 2019;35(6):416-421. doi:10.1017/S0266462319000667
29. Chambers JD, Cangelosi MJ, Neumann PJ. Medicare's use of cost-effectiveness analysis for prevention (but not for treatment). *Health Policy*. 2015;119(2):156-163. doi:10.1016/j.healthpol.2014.11.012
30. Bryan S, Sofaer S, Siegelberg T, Gold M. Has the time come for cost-effectiveness analysis in US health care? *Health Economics, Policy and Law*. 2009;4(4):425-443. doi:10.1017/S1744133109004885
31. Chambers JD, Lord J, Cohen JT, Neumann PJ, Buxton MJ. Illustrating potential efficiency gains from using cost-effectiveness evidence to reallocate medicare expenditures. *Value in Health*. 2013;16(4):629-638. doi:10.1016/j.jval.2013.02.011
32. Nemzoff C, Ruiz F, Chalkidou K, et al. Adaptive health technology assessment to facilitate priority setting in low-and middle-income countries. *BMJ Global Health*. 2021;6(4):4549. doi:10.1136/bmjgh-2020-004549
33. Tantivess S, Chalkidou K, Tritasavit N, Teerawattananon Y. Health Technology

Assessment capacity development in low- and middle-income countries: Experiences from the international units of HITAP and NICE. *F1000Research*. 2017;6. doi:10.12688/f1000research.13180.1

34. Babigumira JB, Jenny AM, Bartlein R, Stergachis A, Garrison LP. Health technology assessment in low- and middle-income countries: A landscape assessment. *Journal of Pharmaceutical Health Services Research*. 2016;7(1):37-42. doi:10.1111/jphs.12120
35. Luz A, Santatiwongchai B, Pattanaphesaj J, Teerawattananon Y. Identifying priority technical and context-specific issues in improving the conduct, reporting and use of health economic evaluation in low- and middle-income countries. *Health Research Policy and Systems*. 2018;16(1):1-12. doi:10.1186/S12961-018-0280-6
36. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *British Medical Journal*. 2004;329(7459):224-227. doi:10.1136/BMJ.329.7459.224
37. National Institute for Health and Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal: 2013. Published 2013. Accessed February 27, 2022. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781
38. NICE Decision Support Unit. Technical Support Documents. Accessed February 27, 2022. https://nicedsu.sites.sheffield.ac.uk/tsds/completed-or-in-progress
39. Noon K, Ulrich A, Painter C, Montgomery S. When The Price Is Right: When, In Practice, Are Confidential Discounts Introduced In The Nice Process? *Value in Health*. 2017;20(9):A695. doi:10.1016/J.JVAL.2017.08.1783
40. *The National Institute for Health and Care Excellence (Constitution and Functions) and the Health and Social Care Information Centre (Functions) Regulations 2013*.; 2013. https://www.legislation.gov.uk/uksi/2013/259/made
41. Tousignant P, Guttmann RD, Hollomby DJ. Transplantation and home hemodialysis: Their cost-effectiveness. *Journal of Chronic Diseases*. 1985;38(7):589-601. doi:10.1016/0021-9681(85)90048-7
42. Williams A. Economics of coronary artery bypass grafting. *British Medical Journal*. 1985;291(6491):326-329. doi:10.1136/BMJ.291.6491.326
43. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: Design, conduct, analysis, and reporting. *British Medical Journal*. 2011;342(7806). doi:10.1136/bmj.d1548
44. Yeo MM. Conducting Economic Evaluations Alongside Randomised Trials: Current

Methodological Issues and Novel Approaches. *Pharmacoeconomics*. 2016;34(5):447-461. doi:10.1007/s40273-015-0371-y

45. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics*. 2006;15(7):677-687. doi:10.1002/hec.1093
46. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life. *Health Economics*. 1998;6(3):217-227. doi:10.1002/(SICI)1099-1050(199705)6:3<217::AID-HEC267>3.0.CO;2-W
47. Pitt M, Monks T, Crowe S, Vasilakis C. Systems modelling and simulation in health service design, delivery and decision making. *BMJ Quality & Safety*. 2016;25(1):38-45. doi:10.1136/BMJQS-2015-004430
48. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics*. 2003;12(10):837-848. doi:10.1002/HEC.770
49. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics*. 2006;15(12):1295-1310. doi:10.1002/HEC.1148
50. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: Selecting the appropriate approach. *Journal of Health Services Research and Policy*. 2004;9(2):110-118. doi:10.1258/135581904322987535
51. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Management Science 1998 1:2*. 1998;1(2):133-140. doi:10.1023/A:1019090401655
52. Sonnenberg FA, Beck JR. Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making*. 1993;13(4):322-338. doi:10.1177/0272989X9301300409
53. Briggs A, Sculpher M. An Introduction to Markov Modelling for Economic Evaluation. *PharmacoEconomics 1998 13:4*. 2012;13(4):397-409. doi:10.2165/00019053-199813040-00003
54. Campbell H, Briggs A, Buxton M, Kim L, Thompson S. The credibility of health economic models for health policy decision-making: The case of population screening for abdominal aortic aneurysm. *Journal of Health Services Research and Policy*. 2007;12(1):11-17. doi:10.1258/135581907779497594
55. Kim LG, Thompson SG. Uncertainty and validation of health economic decision

models. *Health Economics*. 2010;19(1):43-55. doi:10.1002/HEC.1444

56. Claxton K. The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*. 1999;18(3):341-364. doi:10.1016/S0167-6296(98)00039-3

57. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: The consequences of ignoring decision uncertainty. *Health Economics*. 2011;20(2):212-224. doi:10.1002/hec.1586

58. Reinhold T, Brüggenjürgen B, Schlander M, Rosenfeld S, Hessel F, Willich SN. Economic analysis based on multinational studies: Methods for adapting findings to national contexts. *Journal of Public Health*. 2010;18(4):327-335. doi:10.1007/S10389-010-0315-0/FIGURES/4

59. Brereton N, Pennington B, Ekelund M, Akehurst R. A cost-effectiveness analysis of celecoxib compared with diclofenac in the treatment of pain in osteoarthritis (OA) within the Swedish health system using an adaptation of the NICE OA model. *Journal of Medical Economics*. 2014;17(9):677-684. doi:10.3111/13696998.2014.933111

60. Daniel Mullins C, Onwudiwe NC, Branco de Araújo GT, et al. Guidance Document: Global Pharmacoeconomic Model Adaption Strategies. *Value in Health Regional Issues*. 2014;5:7-13. doi:10.1016/J.VHRI.2014.02.006

61. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS Outcomes Model 2: A new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year united kingdom prospective diabetes Study: UKPDS 82. *Diabetologia*. 2013;56(9):1925-1933. doi:10.1007/S00125-013-2940-Y/FIGURES/3

62. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment. *PharmacoEconomics*. 2012;24(4):355-371. doi:10.2165/00019053-200624040-00006

63. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decisionanalytic modeling in health care evaluation. *Value in Health*. 2003;6(1):9-17. doi:10.1046/j.1524-4733.2003.00234.x

64. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. *Medical Decision Making*. 2012;32(5):667-677. doi:10.1177/0272989X12454577

65. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes

modeling practices: a suggested framework. *Value in Health*. 1998;1(2):131-147. doi:10.1046/j.1524-4733.1998.120131.x

66. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics*. 2006;15(7):677-687. doi:10.1002/hec.1093
67. Hanney S, Buxton M, Green C, Coulson D, Raftery J. An assessment of the impact of NHS Health Technology Assessment Programme. *Health Technology Assessment*. 2007;11(53). doi:10.3310/HTA11530
68. Weel C van. Biomedical science matters for people—so its impact should be better assessed. *The Lancet*. 2002;360(9339):1034-1035. doi:10.1016/S0140-6736(02)11175-5
69. Hanney SR, Gonzalez-Block MA, Buxton MJ, Kogan M. The utilisation of health research in policy-making: concepts, examples and methods of assessment. *Health Research Policy and Systems*. 2003;1(1):1-28. doi:10.1186/1478-4505-1-2
70. Buxton M, Hanney S, Jones T. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organization*. 2004;82(10):733-739. doi:10.1590/S0042-96862004001000007
71. Buxton M, Hanney S, Packwood T, Roberts S, Youll P. Assessing benefits from department of health and national health service research & development. *Public Money and Management*. 2000;20(4):29-34. doi:10.1111/1467-9302.00233
72. Hanney SR, Watt A, Jones TH, Metcalf L. Conducting retrospective impact analysis to inform a medical research charity's funding strategies: The case of Asthma UK. *Allergy, Asthma and Clinical Immunology*. 2013;9(1):1-12. doi:10.1186/1710-1492-9-17/TABLES/1
73. Nason E, Janta B, Hastings G, Hanney S, O'driscoll M, Wooding S. *Health Research-Making an Impact: The Economic and Social Benefits of HRB-Funded Research.* Health Research Board; 2008. Accessed February 27, 2022. https://www.hrb.ie/funding/evaluation/evaluation-reports/publication/making-an-impact-the-economic-and-social-benefits-of-hrb-funded-research/returnPage/1/
74. Wooding S, Hanney S, Buxton M, Grant J. *The Returns from Arthritis Research Volume 1: Approach, Analysis and Recommendations*. RAND Corporation; 2004. Accessed February 27, 2022. https://www.rand.org/pubs/monographs/MG251.html. Also available in print form
75. Smit JP, Hessels LK. The production of scientific and societal value in research

evaluation: a review of societal impact assessment methods. *Research Evaluation*. 2021;30(3):323-335. doi:10.1093/RESEVAL/RVAB002

76. Murphy KM, Topel RH. The Economic Value of Medical Research. In: Murphy KM, Topel RH, eds. *Measuring the Gains from Medical Research: An Economic Approach*. University of Chicago Press; 2003.
77. Heidbrink C, Mcguire A, Raikou M. Inferring the Value of Medical Research to the UK. *LSE Health*. Published online 2007. Accessed August 9, 2021. http://eprints.lse.ac.uk/22632/
78. Townsend J, Buxton M. Cost effectiveness scenario analysis for a proposed trial of hormone replacement therapy. *Health Policy*. 1997;39(3):181-194. doi:10.1016/S0168-8510(96)00870-6
79. Guthrie S, Bienkowska-Gibbs T, Manville C, Pollitt A, Kirtley A, Wooding S. The impact of the National Institute for Health Research Health Technology Assessment programme. *Health Technology Assessment*. 2015;19(67):1-292. doi:10.3310/HTA19670
80. Fenwick E, Claxton K, Sculpher M. The value of implementation and the value of information: combined and uneven development. *Medical Decision Making*. 2008;28(1):21-32. doi:10.1177/0272989X07308751
81. Claxton KP, Sculpher MJ. Using Value of Information Analysis to Prioritise Health Research. *PharmacoEconomics 2006 24:11*. 2012;24(11):1055-1068. doi:10.2165/00019053-200624110-00003
82. Sakowsky RA. Disentangling the welfarism/extra-welfarism distinction: Towards a more fine-grained categorization. *Health Economics*. 2021;30(9):2307-2311. doi:10.1002/HEC.4382
83. Birch S, Donaldson C. Valuing the benefits and costs of health care programmes: Where's the "extra" in extra-welfarism? *Social Science and Medicine*. 2003;56(5):1121-1133. doi:10.1016/S0277-9536(02)00101-6
84. McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: What it is and what that means. *PharmacoEconomics*. 2008;26(9):733-744. doi:10.2165/00019053-200826090-00004/FIGURES/3
85. Greenhalgh T, Raftery J, Hanney S, Glover M. Research impact: a narrative review. *BMC Medicine 2016 14:1*. 2016;14(1):1-16. doi:10.1186/S12916-016-0620-8
86. Donovan C, Hanney S. The 'Payback Framework' explained. *Research Evaluation*. 2011;20(3):181-183. doi:10.3152/095820211X13118583635756

87. Buxton M, Hanney S. How can payback from health services research be assessed? *Journal of Health Services Research & Policy*. 1996;1(1):35-43. doi:10.1177/135581969600100107
88. Kuruvilla S, Mays N, Pleasant A, Walt G. Describing the impact of health research: A Research Impact Framework. *BMC Health Services Research*. 2006;6(1):1-18. doi:10.1186/1472-6963-6-134/TABLES/2
89. Panel on Return on Investment in Health Research. *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. Canadian Academy of Health Sciences; 2009. Accessed February 27, 2022. https://cahs-acss.ca/wp-content/uploads/2011/09/ROI_FullReport.pdf
90. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—Explanation and Elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value in Health*. 2013;16(2):231-250. doi:10.1016/J.JVAL.2013.02.002
91. Drummond M. Output measurement for resource allocation decisions in health care. *Oxford Review of Economic Policy*. 1989;5(1):59-74. doi:10.1093/oxrep/5.1.59
92. Alla K, Hall WD, Whiteford HA, Head BW, Meurk CS. How do we define the policy impact of public health research? A systematic review. *Health Research Policy and Systems*. 2017;15(1):1-12. doi:10.1186/S12961-017-0247-Z/TABLES/2
93. Grant J, Buxton MJ. Economic returns to medical research funding. *BMJ Open*. 2018;8:22131. doi:10.1136/bmjopen-2018-022131
94. Mushkin SJ, Landefeld JS. *Biomedical Research: Costs and Benefits*. Ballinger Pub. Co.; 1979.
95. Toole AA. Does public scientific research complement private investment in research and development in the pharmaceutical industry? *Journal of Law and Economics*. 2007;50(1):81-104. doi:10.1086/508314
96. Health Economics Research Group, Office of Health Economics, RAND Europe. *Medical Research: What's It Worth? Estimating the Economic Benefits from Medical Research in the UK*. UK Evaluation Forum; 2008. Accessed February 27, 2022. https://mrc.ukri.org/publications/browse/medical-research-whats-it-worth/
97. Sussex J, Feng Y, Mestre-Ferrandiz J, et al. Quantifying the economic impact of government and charity funding of medical research on private research and development funding in the United Kingdom. *BMC Medicine*. 2016;14(1):1-23. doi:10.1186/S12916-016-0564-Z/FIGURES/6

98. Roback K, Dalal K, Carlsson P. Evaluation of health research: measuring costs and socioeconomic effects. *International journal of preventive medicine*. 2011;2(4):203-215.
99. Lakdawalla DN, Sun EC, Jena AB, Reyes CM, Goldman DP, Philipson TJ. An economic evaluation of the war on cancer. *Journal of Health Economics*. 2010;29(3):333-346. doi:10.1016/j.jhealeco.2010.02.006
100. Access Economics. *Exceptional Returns: The Value of Investing in Health R&D in Australia, Report for the Australian Society for Medical Research*. Australian Society for Medical Research; 2003. Accessed February 27, 2022. https://asmr.org.au/wp-content/uploads/library/Except.pdf
101. Deloitte Access Economics. *Australia's Health and Medical Research Workforce: Expert People Providing Exceptional Returns*. Deloitte Access Economics Pty Ltd; 2016. Accessed February 27, 2022. https://www2.deloitte.com/au/en/pages/economics/articles/australias-health-and-medical-research-workforce.html
102. Deloitte Access Economics. *Returns on NHMRC Funded Research and Development, Report for Australian Society for Medical Research*. Deloitte Access Economics Pty Ltd; 2011. Accessed February 27, 2022. https://asmr.org.au/wp-content/uploads/library/NHMRCReturns.pdf
103. Deloitte Access Economics. *Extrapolated Returns on Investment in NHMRC Medical Research, Report for Australian Society for Medical Research*. Deloitte Access Economics Pty Ltd; 2014. Accessed February 27, 2022. https://asmr.org.au/wp-content/uploads/library/ASMR Deloittee Report_MRFF.pdf
104. Access Economics. *Exceptional Returns: The Value of Investing in Health R&D in Australia II, Report for the Australian Society for Medical Research*. Australian Society for Medical Research; 2008. Accessed February 27, 2022. https://asmr.org.au/wp-content/uploads/library/ExceptII08.pdf
105. Johnston SC, Rootenberg JD, Katrak S, Smith WS, Elkins JS. Effect of a US National Institutes of Health programme of clinical trials on public health and costs. *Lancet*. 2006;367(9519):1319-1327. doi:10.1016/S0140-6736(06)68578-4
106. Stinnett AA, Mullahy J. Net health benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making*. 1998;18(2 SUPPL.). doi:10.1177/0272989x98018002s09
107. Grant J, Cottrell R, Fawcett G, Cluzeau F. Evaluating "payback" on biomedical

research from papers cited in clinical guidelines: applied bibliometric study. *BMJ*. 2000;320(7242):1107-1111. doi:10.1136/BMJ.320.7242.1107

108. Guthrie S, Pollitt A, Hanney S, Grant J. Investigating Time Lags and Attribution in the Translation of Cancer Research: A Case Study Approach. *RAND Health Quarterly*. 2014;4(2):16.
109. Rao X, Hurley TM, Pardey PG. Recalibrating the reported returns to agricultural R&D: what if we all heeded Griliches?*. *Australian Journal of Agricultural and Resource Economics*. 2020;64(3):977-1001. doi:10.1111/1467-8489.12388
110. Kim LG, Thompson SG, Briggs AH, Buxton MJ, Campbell HE. How cost-effective is screening for abdominal aortic aneurysms? *Journal of Medical Screening*. 2007;14(1):46-52. doi:10.1258/096914107780154477
111. McDaid C, Griffin S, Weatherly H, et al. Continuous positive airway pressure devices for the treatment of obstructive sleep apnoea-hypopnoea syndrome: a systematic review and economic analysis. *Health Technology Assessment*. 2009;13(4):1-142. doi:10.3310/HTA13040
112. Scott RAP, Ashton HA, Buxton MJ, et al. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *The Lancet*. 2002;360(9345):1531-1539. doi:10.1016/S0140-6736(02)11522-4
113. Sweeting MJ, collaborators on behalf of the R, Thompson SG, et al. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *British Journal of Surgery*. 2012;99(5):655-665. doi:10.1002/BJS.8707
114. Sweeting MJ, Masconi KL, Jones E, et al. Analysis of clinical benefit, harms, and cost-effectiveness of screening women for abdominal aortic aneurysm. *The Lancet*. 2018;392(10146):487-495. doi:10.1016/S0140-6736(18)31222-4
115. Sonnenberg FA, Beck JR. Markov Models in Medical Decision Making: A Practical Guide. *Medical Decision Making*. 1993;13(4):322-338. doi:10.1177/0272989X9301300409
116. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? *Value in Health*. 2010;13(8):1056-1060. doi:10.1111/J.1524-4733.2010.00775.X
117. Greenhalgh T, Fahy N. Research impact in the community-based health sciences: An analysis of 162 case studies from the 2014 UK Research Excellence Framework. *BMC*

Medicine. 2015;13(1):1-12. doi:10.1186/S12916-015-0467-4/FIGURES/4

118. Searles A, Doran C, Attia J, et al. An approach to measuring and encouraging research translation and research impact. *Health Research Policy and Systems*. 2016;14(1):1-13. doi:10.1186/S12961-016-0131-2/FIGURES/3
119. Emery P, Breedveld FC, Dougados M, Kalden JR. Early referral recommendation for newly diagnosed rheumatoid arthritis: evidence based development of a clinical guide. *Annals of the Rheumatic Diseases*. 2002;61(4):290-297. doi:10.1136/ard.61.4.290
120. Pincus T, Yazici Y, Sokka T, Aletaha D, Smolen JS. Methotrexate as the "anchor drug" for the treatment of early rheumatoid arthritis. *Clinical and Experimental Rheumatology*. 2003;21(5 Suppl 31):S179-85.
121. Albers JMC, Paimela L, Kurki P, et al. Treatment strategy, disease activity, and outcome in four cohorts of patients with early rheumatoid arthritis. *Annals of the Rheumatic Diseases*. 2001;60(5):453-458. doi:10.1136/ARD.60.5.453
122. Unal B, Critchley JA, Capewell S. Explaining the Decline in Coronary Heart Disease Mortality in England and Wales Between 1981 and 2000. *Circulation*. 2004;109:1101-1107. doi:10.1161/01.CIR.0000118498.35499.B2
123. Storheim K, Zwart JA. Musculoskeletal disorders and the Global Burden of Disease study. *Annals of the Rheumatic Diseases*. 2014;73(6):949-950. doi:10.1136/ANNRHEUMDIS-2014-205327
124. Paisley S. Classification of evidence in decision-analytic models of cost-effectiveness: A content analysis of published reports. *International Journal of Technology Assessment in Health Care*. 2010;26(4):458-462. doi:10.1017/S026646231000098X
125. Kaltenthaler E, Tappenden P, Paisley S. Reviewing the Evidence to Inform the Population of Cost-Effectiveness Models within Health Technology Assessments. *Value in Health*. 2013;16(5):830-836. doi:10.1016/J.JVAL.2013.04.009
126. Sharples LD, Clutterbuck-James AL, Glover MJ, et al. Meta-analysis of randomised controlled trials of oral mandibular advancement devices and continuous positive airway pressure for obstructive sleep apnoea-hypopnoea. *Sleep Medicine Reviews*. 2016;27:108-124. doi:10.1016/J.SMRV.2015.05.003
127. Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. *PharmacoEconomics*. 2016;34(6):597-608. doi:10.1007/S40273-015-0372-X
128. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for

technology assessments. *International Journal of Technology Assessment in Health Care*. 2010;26(4):431-435. doi:10.1017/S0266462310000966

129. Rotstein D, Laupacis A. Differences between systematic reviews and health technology assessments: A trade-off between the ideals of scientific rigor and the realities of policy making. *International Journal of Technology Assessment in Health Care*. 2004;20(2):177-183. doi:10.1017/S0266462304000959
130. Thompson SG, Bown MJ, Glover MJ, et al. Screening women aged 65 years or over for abdominal aortic aneurysm: A modelling study and health economic evaluation. *Health Technology Assessment*. 2018;22(43):1-141. doi:10.3310/HTA22430
131. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *PharmacoEconomics*. 2008;26(9):781-798. doi:10.2165/00019053-200826090-00008
132. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Economics*. 1994;3(2):95-104. doi:10.1002/HEC.4730030206
133. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing Structural Uncertainty in Decision Analytic Models: A Review and Application of Methods. *Value in Health*. 2009;12(5):739-749. doi:10.1111/J.1524-4733.2008.00502.X
134. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A Framework for Addressing Structural Uncertainty in Decision Models. *Medical Decision Making*. 2011;31(4):662-674. doi:10.1177/0272989X11406986
135. Barton GR, Briggs AH, Fenwick EAL. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). *Value in Health*. 2008;11(5):886-897. doi:10.1111/J.1524-4733.2008.00358.X
136. Fenwick E, Marshall DA, Levy AR, Nichol G. Using and interpreting cost-effectiveness acceptability curves: An example using data from a trial of management strategies for atrial fibrillation. *BMC Health Services Research*. 2006;6:52. doi:10.1186/1472-6963-6-52
137. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Economics*. 2001;10(8):779-787. doi:10.1002/HEC.635
138. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and care excellence cost-effectiveness threshold. *Health Technology Assessment*. 2015;19(14):1-503. doi:10.3310/hta19140

139. Thompson SG, Ashton HA, Gao L, Scott RAP. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised Multicentre Aneurysm Screening Study. *BMJ*. 2009;338(7710):b2307. doi:10.1136/BMJ.B2307
140. Buxton MJ, Hanney S, Jones TH. Estimating the economic value to societies of the impact of health research: a critical review. *Bulletin of the World Health Organization*. 2004;82(10):733-739. doi:10.1590/S0042-96862004001000007
141. Grant J, Cottrell R, Fawcett G, Cluzeau F. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *BMJ*. 2000;320(7242):1107-1111. doi:10.1136/BMJ.320.7242.1107
142. Jones TH, Hanney S. Tracing the indirect societal impacts of biomedical research: development and piloting of a technique based on citations. *Scientometrics*. 2016;107(3):975-1003. doi:10.1007/S11192-016-1895-4/TABLES/5
143. Jannot AS, Agoritsas T, Gayet-Ageron A, Perneger T V. Citation bias favoring statistically significant studies was present in medical research. *Journal of Clinical Epidemiology*. 2013;66(3):296-301. doi:10.1016/J.JCLINEPI.2012.09.015
144. Urlings MJE, Duyx B, Swaen GMH, Bouter LM, Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. *Journal of Clinical Epidemiology*. 2021;132:71-78. doi:10.1016/J.JCLINEPI.2020.11.019
145. Bornmann L, Haunschild R, Mutz R. Should citations be field-normalized in evaluative bibliometrics? An empirical analysis based on propensity score matching. *Journal of Informetrics*. 2020;14(4):101098. doi:10.1016/J.JOI.2020.101098
146. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*. 2008;22(2):338-342. doi:10.1096/FJ.07-9492LSF
147. Mongeon P, Paul-Hus A. The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*. 2015;106(1):213-228. doi:10.1007/S11192-015-1765-5
148. Yan E, Ding Y. Weighted citation: An indicator of an article's prestige. *Journal of the American Society for Information Science and Technology*. 2010;61(8):1635-1643. doi:10.1002/ASI.21349
149. Colledge L, on behalf of all the Snowball Metrics program partners. *Snowball Metrics Recipe Book*. Edition 2. Elsevier; 2014. Accessed February 27, 2022.

https://www.snowballmetrics.com/wp-content/uploads/snowball-recipe-book_HR.pdf

150. Hanney S, Frame I, Grant J, Buxton M, Young T, Lewison G. Using categorisations of citations when assessing the outcomes from health research. *Scientometrics*. 2005;65(3):357-379. doi:10.1007/S11192-005-0279-Y

151. King's College London and Digital Science. *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies*. HEFCE; 2015. Accessed February 27, 2022. https://www.kcl.ac.uk/policy-institute/research-analysis/nature-scale-beneficiaries-research-impact

152. Sweeting MJ, Marshall J, Glover M, Nasim A, Bown MJ. Evaluating the Cost-Effectiveness of Changes to the Surveillance Intervals in the UK Abdominal Aortic Aneurysm Screening Programme. *Value in Health*. 2021;24(3):369-376. doi:10.1016/J.JVAL.2020.10.015

153. Kim LG, Sweeting MJ, Armer M, Jacomelli J, Nasim A, Harrison SC. Modelling the impact of changes to abdominal aortic aneurysm screening and treatment services in England during the COVID-19 pandemic. *PLOS ONE*. 2021;16(6):e0253327. doi:10.1371/JOURNAL.PONE.0253327

154. Earnshaw JJ, Lees T. Update on Screening for Abdominal Aortic Aneurysm. *European Journal of Vascular and Endovascular Surgery*. 2017;54(1):1-2. doi:10.1016/J.EJVS.2017.04.002

155. Earnshaw JJ. Final Say on Research Into Population Screening for Abdominal Aortic Aneurysm. *Angiology*. 2018;69(5):365-366. doi:10.1177/0003319717705531

156. Björck M, Bown MJ, Choke E, et al. International update on screening for abdominal aortic aneurysms: issues and opportunities. *European Journal of Vascular and Endovascular Surgery*. 2015;49(2):113-115. doi:10.1016/J.EJVS.2014.08.015

157. Mackie A, Latinovic R, Ramsay M, Pebody R. Screening and Immunisation. In: Moore M, ed. *Annual Report of the Chief Medical Officer 2015. On the State of the Public's Health. Babay Boomers: Fit for the Future*. Department of Health and Social Care; 2015:154. Accessed February 27, 2022. https://www.gov.uk/government/publications/cmo-annual-report-2015-health-of-the-baby-boomer-generation

158. Nair N, Kvizhinadze G, Jones GT, et al. Health gains, costs and cost-effectiveness of a population-based screening programme for abdominal aortic aneurysms. *British Journal of Surgery*. 2019;106(8):1043-1054. doi:10.1002/BJS.11169

159. Daroudi R, Shafe O, Moosavi J, et al. Cost-effectiveness of a population-based AAA screening program for men over 65 years old in Iran. *Cost Effectiveness and Resource Allocation*. 2021;19(1):1-10. doi:10.1186/S12962-021-00283-7/FIGURES/3
160. Zarrouk M, Lundqvist A, Holst J, Troëng T, Gottsäter A. Cost-effectiveness of Screening for Abdominal Aortic Aneurysm in Combination with Medical Intervention in Patients with Small Aneurysms. *European Journal of Vascular and Endovascular Surgery*. 2016;51:766-773. doi:10.1016/j.ejvs.2015.12.048
161. Reile R, Võrno T, Kals J, Ilves P, Kiivet RA. The Cost-Effectiveness of Abdominal Aortic Aneurysm Screening in Estonia. *Value in Health Regional Issues*. 2020;22:1-6. doi:10.1016/J.VHRI.2019.08.477
162. Rossi SH, Klatte T, Usher-Smith JA, et al. A Decision Analysis Evaluating Screening for Kidney Cancer Using Focused Renal Ultrasound. *European Urology Focus*. 2021;7(2):407-419. doi:10.1016/J.EUF.2019.09.002
163. SBU. *Screening För Bukaortaaneurysm Bland 65-Åriga Män: Hälsoekonomisk Analys Bilaga 2*. Statens beredning för medicinsk och social utvärdering (SBU); 2015. Accessed February 27, 2022. www.sbu.se/244
164. Wanhainen A, Hultgren R, Linné A, et al. Outcome of the Swedish Nationwide Abdominal Aortic Aneurysm Screening Program. *Circulation*. 2016;134(16):1141-1148. doi:10.1161/CIRCULATIONAHA.116.022305
165. Kapila V, Jetty P, Wooster D, Vucemilo V, Dubois L. Screening for abdominal aortic aneurysms in Canada: 2020 review and position statement of the Canadian Society for Vascular Surgery. *Canadian Journal of Surgery*. 2021;64(5):E461-E466. doi:10.1503/CJS.009120
166. Singh H, Bell N, Dickinson JA, et al. Recommendations on screening for abdominal aortic aneurysm in primary care. *CMAJ*. 2017;189(36):E1137-E1145. doi:10.1503/CMAJ.170118/-/DC1
167. Debus E, Heidemann F, Gross-Fengels W, et al. *S3-Leitlinie Zu Screening, Diagnostik, Therapie Und Nachsorge Des Bauchaortenaneurysmas*. AWMF online. Das Portal der wissenschaftlichen Medizin; 2018. Accessed February 27, 2022. https://www.awmf.org/uploads/tx_szleitlinien/004-014l__S3_Bauchaortenaneurysma_2018-08.pdf
168. Lahoz C, Gracia CE, García LR, et al. SEA-SEACV 2015: Guía para el diagnóstico y tratamiento del aneurisma de aorta abdominal. *Clínica e Investigación en Arteriosclerosis*. 2016;28:1-49. doi:10.1016/S0214-9168(16)30026-2

169. Kozak M, Polanec B, Gasparini M, et al. Smernice za obravnavo bolnikov z anevrizmo trebušne aorte. *Zdrav Vestn (Slovenian Medical Journal)*. 2015;84:583-590.
170. Wanhainen A, Verzini F, Van Herzeele I, et al. Editor's Choice - European Society for Vascular Surgery (ESVS) 2019 Clinical Practice Guidelines on the Management of Abdominal Aorto-iliac Artery Aneurysms. *European Journal of Vascular and Endovascular Surgery*. 2019;57(1):8-93. doi:10.1016/J.EJVS.2018.09.020
171. Zierler RE, Jordan WD, Lal BK, et al. The Society for Vascular Surgery practice guidelines on follow-up after vascular surgery arterial procedures. *Journal of Vascular Surgery*. 2018;68(1):256-284. doi:10.1016/J.JVS.2018.04.018
172. Chaikof EL, Dalman RL, Eskandari MK, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *Journal of Vascular Surgery*. 2018;67(1):2-77.e2. doi:10.1016/J.JVS.2017.10.044
173. National Institute for Health and Care Excellence. *Abdominal Aortic Aneurysm: Diagnosis and Management Evidence Review D: Monitoring for Abdominal Aortic Aneurysm Expansion and Risk of Rupture NICE Guideline NG156*.; 2020. Accessed February 27, 2022. https://www.nice.org.uk/guidance/ng156/evidence/d-monitoring-for-abdominal-aortic-aneurysm-expansion-and-risk-of-rupture-pdf-7087991153
174. National Institute for Health and Care Excellence. *Obstructive Sleep Apnoea/ Hypopnoea Syndrome and Obesity Hypoventilation Syndrome in over 16s. Evidence Review G: Oral Devices. NG202.*; 2021. Accessed January 7, 2022. https://www.nice.org.uk/guidance/ng202/evidence/g-oral-devices-pdf-9204444691
175. Weng SJ, Tsai MC, Tsai Y Te, et al. Improving the Efficiency of an Emergency Department Based on Activity-Relationship Diagram and Radio Frequency Identification Technology. *International Journal of Environmental Research and Public Health 2019, Vol 16, Page 4478*. 2019;16(22):4478. doi:10.3390/IJERPH16224478
176. Marrero WJ, Lavieri MS, Sussman JB. Optimal cholesterol treatment plans and genetic testing strategies for cardiovascular diseases. *Health Care Management Science*. 2021;24(1):1-25. doi:10.1007/S10729-020-09537-X/METRICS
177. Apostolopolous Y. Harnessing Complex Systems. In: Apostolopoulos Y, Lich KH, Lemke MK, eds. *Complex Systems and Population Health*. Oxford University Press; 2020:283-298.
178. Improta G, Guizzi G, Ricciardi C, et al. Agile Six Sigma in Healthcare: Case Study at Santobono Pediatric Hospital. *International Journal of Environmental Research and*

Public Health 2020, Vol 17, Page 1052. 2020;17(3):1052. doi:10.3390/IJERPH17031052

179. Mohammed MA, Kadhim Mohsin S, Jasim Mohammed S, Mohammed maysoonamohammed MA. The Effectiveness of Using Discrete Event Simulation to Optimize the Quality of Service of Outpatient in Iraq: A Case Study. *Iraqi Journal of Industrial Research*. 2021;8(1):40-49. doi:10.53523/IJOIRVOL8I1ID1
180. Tamburis O. Bridging the gap between process mining and des modeling in the healthcare domain. *2019 7th E-Health and Bioengineering Conference, EHB 2019*. Published online November 1, 2019. doi:10.1109/EHB47216.2019.8969912
181. Tamburis O, Esposito C. Process mining as support to simulation modeling: A hospital-based case study. *Simulation Modelling Practice and Theory*. 2020;104:102149. doi:10.1016/J.SIMPAT.2020.102149
182. Smith JBE, Channon K, Kiparoglou V, Forbes JF, Gray AM. A macroeconomic assessment of the impact of medical research expenditure: A case study of NIHR Biomedical Research Centres. *PLOS ONE*. 2019;14(4):e0214361. doi:10.1371/JOURNAL.PONE.0214361
183. Meghea CI, Montgomery BW, Ellington R, et al. An NIH investment in health equity - the economic impact of the Flint Center for Health Equity Solutions. *BMC Public Health*. 2021;21(1):1-10. doi:10.1186/S12889-021-11795-5/TABLES/4
184. Hanna CR, Boyd KA, Jones RJ. Evaluating cancer research impact: lessons and examples from existing reviews on approaches to research impact assessment. *Health Research Policy and Systems*. 2021;19(1):1-45. doi:10.1186/S12961-020-00658-X/FIGURES/3
185. Gomes D, Stavropoulou C. The impact generated by publicly and charity-funded research in the United Kingdom: A systematic literature review. *Health Research Policy and Systems*. 2019;17(1):1-11. doi:10.1186/S12961-019-0425-2/TABLES/2
186. Hanney S, Greenhalgh T, Blatch-Jones A, Glover M, Raftery J. The impact on healthcare, policy and practice from 36 multi-project research programmes: Findings from two reviews. *Health Research Policy and Systems*. 2017;15(1):1-21. doi:10.1186/S12961-017-0191-Y/TABLES/5
187. Coombs SK, Meijer I. Towards Evaluating the Research Impact made by Universities of Applied Sciences. *Science and Public Policy*. 2021;48(2):226-234. doi:10.1093/SCIPOL/SCAB009
188. Millar R, Morton A, Bufali MV, et al. Assessing the performance of health technology

assessment (HTA) agencies: developing a multi-country, multi-stakeholder, and multi-dimensional framework to explore mechanisms of impact. *Cost Effectiveness and Resource Allocation*. 2021;19(1):1-14. doi:10.1186/S12962-021-00290-8/TABLES/1

189. World Health Organization, Bertram M, Dhaene G, Tan-Torres Edejer T. *Institutionalizing Health Technology Assessment Mechanisms: A How to Guide*. World Health Organization; 2021. Accessed February 27, 2022. https://apps.who.int/iris/handle/10665/340722
190. Buffardi AL, Sharp S, Hadley S, Archer RA. *Measuring Evidence-Informed Decision-Making Processes in Low-and Middle-Income Countries: Briefing Note*. Overseas Development Institute; 2020. Accessed February 27, 2022. https://odi.org/en/publications/measuring-evidence-informed-decision-making-processes-in-low-and-middle-income-countries/
191. Grant J, Montague E, Buxton MJ, Glover MJ. Bang for your buck: the return from public investments in biomedical and health research - On Health. *BMC*. https://blogs.biomedcentral.com/on-health/2018/01/16/bang-for-your-buck-the-return-from-public-investments-in-biomedical-and-health-research/. Published 2018. Accessed February 27, 2022.
192. Shaiki-Lesko R. Cancer Research ROI. Report quantifies the returns from the U.K.'s investments in oncology. *The Scientist*. https://www.the-scientist.com/the-nutshell/cancer-research-roi-37328. Published 2014. Accessed February 27, 2022.
193. Hawkes N. Cancer research provides a good economic return as well as health gains, report says. *BMJ*. 2014;348. doi:10.1136/BMJ.G3973
194. All Party Parliamentary Group on Medical Research. *A Healthy Future for UK Medical Research*. Medical Research Council; 2014. Accessed January 7, 2022. https://mrc.ukri.org/publications/browse/a-healthy-future-for-uk-medical-research/
195. The Academy of Medical Sciences, Cancer Research UK, Department of Health, Wellcome Trust. *Medical Research: What's It Worth? Estimating the Economic Benefits of Cancer-Related Research in the UK.* Wellcome Trust; 2014. Accessed January 7, 2022. https://acmedsci.ac.uk/file-download/35847-53d2235324225.pdf
196. Wellcome Trust, National Institute for Health Research, The Academy of Medical Sciences, Medical Research Council, Arthritis Research UK. *Medical Research: What's It Worth?* Wellcome; 2018. Accessed January 7, 2022. https://wellcome.org/sites/default/files/whats-it-worth-musculoskeletal-disease-research-januar-2018.pdf

197. MRC. Medical Research Council Annual Report and. Published online 2017. Accessed January 7, 2022. www.gov.uk/government/publications
198. Neyt M, Christiaens Universiteit Gent T, Centrum voor Farmacotherapeutische Informatie B, et al. *Publicly Funded Practice-Oriented Clinical Trials*. Belgian Health Care Knowledge Centre (KCE); 2015. Accessed January 7, 2022. https://kce.fgov.be/sites/default/files/atoms/files/KCE_246_Public_funded_clinical_trials_Report.pdf
199. Cancer Research UK. Written evidence submitted by Cancer Research UK (COV0092) Summary. Published online 2020. Accessed January 7, 2022. https://committees.parliament.uk/writtenevidence/4128/pdf/
200. Alan Turing Institute. Science and Technology Select Committee. Life Sciences Industrial Strategy: Who's driving the bus? Oral and Written evidence Contents. Published online 2017. Accessed January 7, 2022. https://www.parliament.uk/globalassets/documents/lords-committees/science-technology/life-sciences-industrial-strategy/Life-sciences-industrial-strat-evidence.pdf
201. Brunel University London. Report finds that investing in cancer research boosts economy and health.https://www.brunel.ac.uk/news-and-events/news/articles/Report-finds-that-investing-in-cancer-research-boosts-economy-and-health. Published 2014. Accessed January 9, 2022.
202. The Academy of Medical Sciences. Outstanding return on investment in medical research.https://acmedsci.ac.uk/more/news/outstanding-return-on-investment-in-medical-research. Published 2018. Accessed January 9, 2022.
203. Meghea CI, Montgomery BW, Ellington R, et al. An NIH investment in health equity - the economic impact of the Flint Center for Health Equity Solutions. *BMC Public Health*. 2021;21(1):1-10. doi:10.1186/S12889-021-11795-5/TABLES/4
204. Craig J, Avila AC, Dale V, Bloor K, Hex N. *Estimating the Economic Value of NIHR Biomedical Research Centres and Units*.; 2020.
205. Hanna CR. Cancer clinical trials: identifying & testing approaches to impact evaluation. Published online 2021.
206. Boehler CEH, Lord J. Mind the gap! A multilevel analysis of factors related to variation in published cost-effectiveness estimates within and between countries. *Medical Decision Making*. 2016;36(1):31-47. doi:10.1177/0272989X15579173/ASSET/IMAGES/LARGE/10.1177_0272989X15579173-FIG2.JPEG

207. Goeree R, Burke N, O'Reilly D, Manca A, Blackhouse G, Tarride JE. Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. *https://doi.org/101185/030079906X167327*. 2007;23(4):671-682. doi:10.1185/030079906X167327
208. Drummond M, Barbieri M, Cook J, et al. Transferability of Economic Evaluations Across Jurisdictions: ISPOR Good Research Practices Task Force Report. *Value in Health*. 2009;12(4):409-418. doi:10.1111/J.1524-4733.2008.00489.X
209. Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. *PharmacoEconomics*. 2004;22(13):857-876. doi:10.2165/00019053-200422130-00004/FIGURES/TAB4
210. Hanney SR, Castle-Clarke S, Grant J, et al. How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice. *Health Research Policy and Systems*. 2015;13(1). doi:10.1186/1478-4505-13-1
211. Nason E, Curran B, Hanney S, et al. Research Evaluation. *Research Evaluation*. 2011;20(3):193-200. doi:10.3152/095820211X12941371876823
212. Nicholson JM, Mordaunt M, Lopez P, et al. scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*. 2021;2(3):882-898. doi:10.1162/QSS_A_00146
213. Rife S, Rosati D, Nicholson JM. scite: The next generation of citations. *Learned Publishing*. 2021;34(3):454-456. doi:10.1002/leap.1379
214. National Institute for Health and Care Excellence (NICE). NICE publishes new combined methods and processes manual and topic selection manual for its health technology evaluation programmes.https://www.nice.org.uk/news/article/nice-publishes-new-combined-methods-and-processes-manual-and-topic-selection-manual-for-its-health-technology-evaluation-programmes. Published 2022. Accessed February 23, 2022.
215. European Network for Health Technology Assessment (EUnetHTA). Accessed February 27, 2022. https://www.eunethta.eu/
216. THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION. *REGULATION (EU) 2021/2282 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Health Technology Assessment and Amending Directive 2011/24/EU (Text with EEA Relevance). EUR-Lex - 32021R2282 - EN - EUR-Lex.*;

2021. Accessed February 27, 2022. https://eur-lex.europa.eu/eli/reg/2021/2282/oj

217. Morris ZS, wooding S, Grant J. The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine*. 2011;104(12):510-520. doi:10.1258/jrsm.2011.110180

218. Hanney SR, Wooding S, Sussex J, Grant J. From COVID-19 research to vaccine application: Why might it take 17 months not 17 years and what are the wider lessons? *Health Research Policy and Systems*. 2020;18(1):1-10. doi:10.1186/S12961-020-00571-3/FIGURES/1

219. Townsend J, Buxton M, Harper G. Prioritisation of health technology assessment. The PATHS model: methods and case studies. *Health Technology Assessment*. 2003;7(20). doi:10.3310/HTA7200

APPENDIX A: Paper 1

Paper 1: NHS NAAASP

Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England. Br J Surg. 2014;101(8):976-982. https://doi.org/10.1002/bjs.9528

Original article

Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England

M. J. Glover[1], L. G. Kim[2], M. J. Sweeting[3], S. G. Thompson[3] and M. J. Buxton[1]

[1]Health Economics Research Group, Brunel University, and [2]Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, and [3]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
Correspondence to: Mr M. J. Glover, Health Economics Research Group, Brunel University, Uxbridge UB8 3PH, UK (e-mail: Matthew.Glover@brunel.ac.uk)

Background: **Implementation of the National Health Service abdominal aortic aneurysm (AAA) screening programme (NAAASP) for men aged 65 years began in England in 2009. An important element of the evidence base supporting its introduction was the economic modelling of the long-term cost-effectiveness of screening, which was based mainly on 4-year follow-up data from the Multicentre Aneurysm Screening Study (MASS) randomized trial. Concern has been expressed about whether this conclusion of cost-effectiveness still holds, given the early performance parameters, particularly the lower prevalence of AAA observed in NAAASP.**

Methods: **The existing published model was adjusted and updated to reflect the current best evidence. It was recalibrated to mirror the 10-year follow-up data from MASS; the main cost parameters were re-estimated to reflect current practice; and more robust estimates of AAA growth and rupture rates from recent meta-analyses were incorporated, as were key parameters as observed in NAAASP (attendance rates, AAA prevalence and size distributions).**

Results: **The revised and updated model produced estimates of the long-term incremental cost-effectiveness of £5758 (95 per cent confidence interval £4285 to £7410) per life-year gained, or £7370 (£5467 to £9443) per quality-adjusted life-year (QALY) gained.**

Conclusion: **Although the updated parameters, particularly the increased costs and lower AAA prevalence, have increased the cost per QALY, the latest modelling provides evidence that AAA screening as now being implemented in England is still highly cost-effective.**

Paper accepted 7 March 2014
Published online 27 May 2014 in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.9528

Introduction

The UK Multicentre Aneurysm Screening Study (MASS) investigated the effects of offering population screening for abdominal aortic aneurysm (AAA) to men aged 65–74 years. The results of this randomized trial[1], first reported at 4 years of follow-up in 2002, demonstrated that invitation to a one-time ultrasound screen and follow-up of identified aneurysms was effective in reducing AAA-related mortality. This clinical finding has been confirmed by longer-term follow-up from MASS[2–4], and reinforced by systematic reviews[5,6] of evidence including other relevant trials. Based on the initial MASS results it was evident that screening in the context of the UK was likely to be cost-effective in the long-term[7]. This expectation was confirmed by a formal model that extrapolated from the 4-year follow-up data to estimate the long-term incremental cost per quality-adjusted life-year (QALY) for a screening programme of 65-year-old men, using the same screening methods and rescanning intervals for detected aneurysms as in MASS[8]. This estimated the incremental cost per QALY gained for those invited to screening compared with those not invited as £2970 (95 per cent uncertainty interval £2030 to £5430).

In the light of this clinical and cost-effectiveness evidence, and a positive review of all its criteria for a new screening programme, the UK National Screening Committee recommended that a National Health Service (NHS) AAA screening programme (NAAASP) be introduced. Phased implementation began in March 2009 with the aim to cover the whole of England by March 2013[9,10]. Implementation is also under way in Wales, Scotland and Northern Ireland.

Early information from the NAAASP is now available, and it has been noted particularly that the prevalence of AAA at screening is considerably lower than that found in MASS (1·5 per cent compared with 4·9 per cent for MASS)[1,10]. This paper re-estimates the cost-effectiveness of AAA screening as operationalized in England using the most up-to-date available data. The changes to the model reflect: a recalibration to take account of the 10-year follow-up of MASS, using individual patient data; incorporation of updated cost parameters reflecting the current costs of screening, rescans and procedures, including allowance for the introduction of elective endovascular aneurysm repair (EVAR); the use of more robust estimates of AAA growth and rupture rates based on recent meta-analyses[11,12] of individual patient data; and key parameters observed in NAAASP to date (attendance rates, AAA prevalence and aortic size distribution).

Methods

Original model

This re-estimation of the long-term cost-effectiveness of offering AAA screening used the cost-effectiveness model reported in 2007[8]. The underlying Markov model structure is shown in *Fig. 1* and remained unchanged in this reanalysis. The two populations (those invited to AAA screening and those not invited) are modelled using 3-month cycles; each arrow in *Fig. 1* represents a possible transition. The original model incorporated information from a range of sources to chart the detection, growth and treatment of AAAs over time for these populations, using the 4-year follow-up data from MASS as its prime source. It allowed estimation of 30-year costs and benefits of a programme offering a one-off screen to men aged 65 years with repeat scanning annually for aneurysms with a diameter of 3·0–4·4 cm (small AAA) and every 3 months for those with a diameter of 4·5–5·4 cm (medium AAA). Men with aneurysms over 5·4 cm (large AAA) would be referred for consideration for elective surgery. The model adopted an NHS perspective of costs.

Revalidation and recalibration

The original model had been validated against the 4-year MASS data and shown to perform satisfactorily[13]. Using the longer 10-year follow-up data reported for MASS[3], a revalidation exercise was undertaken to assess how well the model predicted the longer-term observed data and to inform recalibration where necessary. Numbers of key events and cost-effectiveness (at 2008–2009 prices) observed in the trial were compared with results from the model.

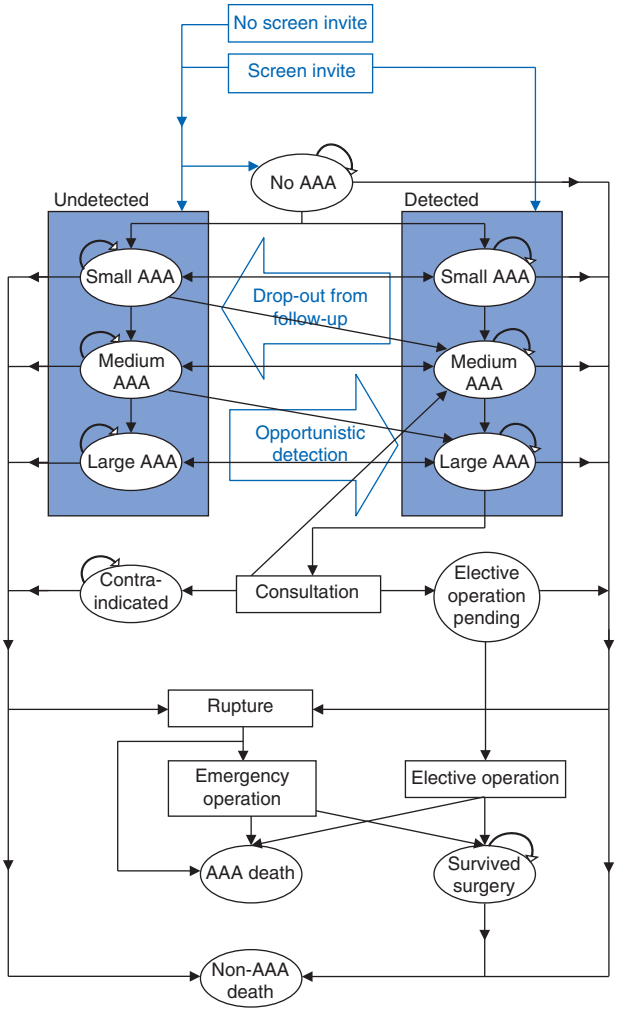


Fig. 1 Markov model structure. AAA, abdominal aortic aneurysm. Reproduced from Kim *et al.*[8], with permission from *Journal of Medical Screening*

To account for any emerging time trends in observed parameters, regression methods were used to derive time-dependent transition probabilities. Based on MASS, 10-year data probabilities were estimated for each 3-monthly cycle, determining transitions between states in the model. Recalibrations of parameter estimates for the rate of opportunistic detection and the rupture rate in large undetected AAAs were also carried out. These parameters cannot be estimated directly from MASS data; hence estimates were chosen to fit the observed data, with a focus on calibration to reflect best the incremental cost-effectiveness ratio (ICER) at 10 years based on observed follow-up. Rates were adjusted to minimize disparity in

Table 1 Unit costs: original estimates from the Multicentre Aneurysm Screening Study, costs inflated to 2010–2011 prices, re-estimated unit costs, cost distributions applied in probabilistic sensitivity analysis, and source

Cost component	Original cost 2000–2001 (£)	MASS cost inflated to 2010–2011 (£)	Re-estimated unit cost (£)	Distribution*	Source
Invitation to screen	1·31	1·84	1·70	Normal(1·7, 0·17)†	NAAASP
Cost of first scan	19·08	26·80	32·20	Normal(32·2, 3·22)†	NAAASP
Surveillance scan	46·04	64·67	68·00	Normal(68·0, 6·80)†	NAAASP
Presurgical assessment	309·88	435·25	435·25	Normal(435·25, 87·05)‡	MASS
Elective repair	6909·00	9704·24	12 806·21	Normal(12 806, 2561)‡	Thompson *et al.*[14]
Emergency repair	11 176·00	15 697·59	19 984·75	Normal(19 985, 3996)‡	Thompson *et al.*[14]

*Normal(μ, σ); standard deviation (σ) †10 per cent and ‡20 per cent of point estimate. MASS, Multicentre Aneurysm Screening Study; NAAASP, National Health Service abdominal aortic aneurysm screening programme.

the modelled and observed differences between arms in key events. A previously published *Health Technology Assessment* monograph[14] deals with this process more comprehensively.

Re-estimation of unit costs

Following the model calibration, input parameters were updated to reflect contemporary costs. The unit cost estimates used in the original modelling related to the costs of screening as undertaken in MASS, and to contemporaneous estimates of the costs of elective and emergency procedures[7]. They were originally estimated at 2000–2001 prices, and in subsequent analyses were simply uplifted to account for general health service inflation. In this updated analysis, costs have been re-estimated and are presented at 2010–2011 price levels. Unit cost data for the screening itself were obtained from NAAASP[14]. Data from MASS[7], the EVAR-1 trial[15] and the National Vascular Database[16] were used to re-estimate the cost of surgical procedures. *Table 1* shows the original aneurysm repair costs, together with the updated unit costs. A fuller account of this re-estimation has been published elsewhere[14].

Clinical data

The majority of probabilistic parameters that determine transitions between states in the Markov model have been updated using the 10-year follow-up data from MASS[3] (*Table 2*). The postcalibration model was also updated to reflect available data from the current NAAASP. Data for attendance rates at screening (75 per cent *versus* 80 per cent in MASS), AAA prevalence (1·5 per cent *versus* 4·9 per cent in MASS) and the size distribution of aneurysms at initial screening (similar in NAAASP and MASS)[10] were incorporated (*Table 2*). Sensitivity analysis around the 30-day surgical mortality rate was also conducted. The mortality rate after elective intervention for a screen-detected AAA observed in the NAAASP was lower (1·6 per cent *versus* 3·0 per cent in MASS), but based on few deaths, so it was deemed inappropriate to use it in the base case. Given the trend of an observed fall in the prevalence rate, a threshold analysis was also conducted to estimate the rate at which the modelling suggests the ICER would rise above £20 000 per QALY.

Growth and rupture rate estimates

The postcalibration model also included improved estimates of aneurysm growth and rupture rates which were derived from the meta-analyses of individual patient data from 18 longitudinal studies of AAA screening surveillance programmes, undertaken as part of the RESCAN Collaboration[11]. The statistical methods used in these meta-analyses have been described elsewhere[11,19], as has their incorporation into the modelling[14].

Implementation of the model

As before, the model was implemented in Microsoft® Excel (Microsoft, San Diego, California, USA), and a 30-year time horizon was adopted (essentially constituting a lifetime for the 65-year-old men considered). Long-term cost and life-years accrued in populations invited to, and not invited to, screening are the outcomes of interest, both discounted at 3·5 per cent per annum. As in previous versions of the modelling, QALYs are estimated by adjusting life-year estimates by EQ-5D™ (EuroQol Group, Rotterdam, The Netherlands) utility values for UK-relevant population age norms[20]. No further adjustment was made, based on the lack of differences in quality of life of those with an AAA[1]. Age-specific death rates from causes other than AAA were taken from UK national statistics[18].

The results are presented as an ICER of invitation to the screening programme compared with no invitation to screening. Probabilistic sensitivity analysis was undertaken

Table 2 Clinical parameters: point estimate used in the model, distribution applied in probabilistic sensitivity analysis, and source

	Estimate	Distribution*	Source
Proportion reinvited to screening	0·1360	Beta(4602, 29 237)	MASS
Prevalence of AAA at first screen			
Attenders	0·0151	Beta(1619, 105 432)	NAAASP
Non-attenders	0·0151	Beta(1619, 105 432)	NAAASP
Non-visualized AAA	0·0151	Beta(1619, 105 432)	NAAASP
Proportion of scans non-visualized	0·0121	Beta(329, 26 818)	MASS
Proportion of screen-invited attending	0·750	Beta(93 170, 31 022)	NAAASP
Proportion of small AAAs at first screen	0·789	Dirichlet(1278, 193, 148)	NAAASP
Proportion of medium AAAs at first screen	0·119		NAAASP
Proportion of large AAAs at first screen	0·091		NAAASP
Transition probabilities (3-monthly)			
Grow from no AAA to small AAA	0·00207	Gamma(27, $7{\cdot}66 \times 10^{-5}$)	Scott *et al.*[17]
Grow from small to medium AAA	TDTP‡	Multiplier ~ Normal(1, 0·1)	RESCAN
Grow from medium to large AAA	TDTP§		RESCAN
Probability of drop-out from surveillance	0·0142	Gamma(330, $4{\cdot}34 \times 10^{-5}$)	MASS
Rupture			
No AAA	0	n.a.	Assumption
Small AAA	TDTP¶	Multiplier ~ Normal(1, 0·35)	RESCAN
Medium AAA	TDTP#		RESCAN
Detected large AAA	0·0125	Gamma(23, 0·00055)	MASS
Undetected large AAA†	0·0282	n.a.	Calibrated
Contraindicated for surgery	0·0282	Gamma(19, 0·0015)	MASS
Opportunistic detection	0·0114	n.a.	Calibrated
Emergency surgery after rupture	0·368	Beta(193, 331)	MASS
Death after emergency surgery	0·342	Beta(66, 127)	MASS
Proportion of large AAAs having surgery	0·681	Dirichlet(481, 156, 69)	MASS
Proportion of large AAAs returned to screening	0·221		MASS
Proportion of large AAAs contraindicated for elective surgery	0·0977		MASS
Death after elective surgery			
Screen-detected AAA	0·0298	Beta(15, 503)	MASS
Opportunistically detected AAA	0·0717	Beta(18, 251)	MASS
All-cause mortality			
Contraindicated for surgery	0·0599	Gamma(41, 0·0015)	MASS
Age-specific	Age-specific	n.a.	Office for National Statistics[18]

*Beta(α, β); Gamma(α, β); Dirichlet($\alpha_1 \ldots \alpha_k$); Normal(μ, σ). †Cannot be observed directly; value chosen during recalibration exercise. ‡Mean 0·016; §mean 0·077; ¶mean 0·00076; #mean 0·0064. MASS, Multicentre Aneurysm Screening Study; AAA, abdominal aortic aneurysm; NAAASP, National Health Service abdominal aortic aneurysm screening programme; TDTP, time-dependent transition probability; RESCAN, RESCAN Collaboration; n.a., not available.

to allow for parameter uncertainty, providing 1000 simulated ICER values. The distributions used for the uncertainty around the point estimate of each variable are detailed in *Tables 1* and *2*. For the updated time-dependent growth and rupture rates, a normally distributed multiplier (with mean 1 and based on a conservative approximation of the standard deviation from the mean of the pooled rates) was defined and sampled from, in order to increase or decrease all growth or rupture rates over time by a constant factor.

Results

The revalidation process showed that the original model did not perform particularly well in predicting the observed MASS 10-year data. There were a number of discrepancies that together led to a substantial difference in the estimate of the 10-year ICER (*Table 3*). Recalibration attempted to minimize the discrepancy in the estimated ICER. The recalibrated model predicted a 10-year ICER of £8900, compared with an ICER based on the 10-year observed data of £7600 per life-year.

The updated 2010–2011 costs for screening and rescans were considerably higher than the 2000–2001 figures originally derived from MASS (*Table 1*). Although this increase reflects general health service inflation, most of these specific costs have increased more rapidly. For example, the cost of elective repair now reflects the proportion of cases in which EVAR is used, leading to a cost that was 32 per cent higher than the inflated value

Table 3 Abdominal aortic aneurysm screening model: validation and recalibration of results using original cost estimates inflated to 2008–2009 prices for consistency

	Observed in MASS*	Original model†	Model after recalibration to MASS 10-year follow-up data‡
Control group			
Elective operations	226	256	213
Emergency operations	141	140	168
AAA deaths	296	305	385
Non-AAA deaths	10 185	10 139	10 148
Life-years (mean)	7·509	7·291	7·282
Mean cost (£)	108	118	124
Invited group			
Elective operations	552	607	539
Emergency operations	62	88	97
AAA deaths	155	202	248
Non-AAA deaths	10 119	10 185	10 189
Mean life-years	7·523	7·297	7·293
Mean cost (£)	208	233	225
Difference between arms			
Elective operations	326	351	326
Emergency operations	−79	−52	−71
AAA deaths	−141	−103	−137
Non-AAA deaths	−66	46	41
Mean difference in life-years	0·013	0·006	0·011
Mean difference in cost (£)	100	115	101
ICER (£)			
Life-years	7600	18 000	8900
QALYs	9700	23 000	11 400

*Key events and cost-effectiveness observed in Multicentre Aneurysm Screening Study (MASS) at 10-year follow-up. †Key events and cost-effectiveness results of modelling, using time-constant parameter estimates from MASS 10-year follow-up. ‡Key events and cost-effectiveness results of modelling, with time-dependent parameter estimates from MASS 10-year follow-up and after recalibration exercise. AAA, abdominal aortic aneurysm; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year (adjusted using population norms).

Table 4 Abdominal aortic aneurysm screening model: 30-year cost-effectiveness results at 2010–2011 prices for the current National Health Service abdominal aortic aneurysm screening programme

	Control group	Invited group	Difference
Life-years†	12·719	12·727	0·0084
QALYs†	9·921	9·928	0·0067
Costs (£)	269	316	47
ICER (£)‡			
Life-years		5758 (4285, 7410)	
QALYs		7370 (5467, 9443)	

Values in parentheses are 95 per cent confidence intervals. Modelling after recalibration, incorporating Multicentre Aneurysm Screening Study (MASS) 10-year follow-up data, growth and rupture rates from meta-analysis of patient-level data, National Health Service abdominal aortic aneurysm screening programme (NAAASP) data on attendance, prevalence and abdominal aortic aneurysm size at initial screen and updated costs. †Life-years and costs discounted at 3·5 per cent. ‡Estimated from the mean of incremental cost-effectiveness ratios (ICERs) produced by 1000 probabilistic sensitivity analysis iterations. QALY, quality-adjusted life-year.

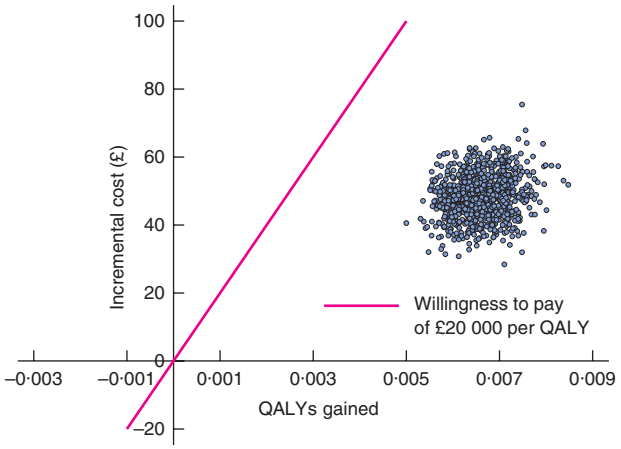


Fig. 2 National Health Service abdominal aortic aneurysm screening programme (NAAASP) cost-effectiveness estimates (30 years); 1000 probabilistic sensitivity analysis iterations. QALY, quality-adjusted life-year

of the original estimate. The estimate for an emergency repair was also 27 per cent higher.

The new estimates of life-years, costs and cost-effectiveness results, over a 30-year time horizon, for an AAA screening programme are shown in *Table 4*. The ICER is now £5758 (95 per cent confidence interval £4285 to £7410) per life-year gained and £7370 (£5467 to £9443) per QALY gained.

When presented on the cost-effectiveness plane (*Fig. 2*), the 1000 iterations of the probabilistic sensitivity analysis show that, in all cases, the intervention provides additional QALYs but costs more. The figure demonstrates the low level of remaining uncertainty and that all estimates fall below the £20 000 threshold, as used by the National Institute for Health and Care Excellence (NICE)[21]. Furthermore, for any threshold value of a QALY over £10 000, there is at least a 99 per cent probability that the programme is cost-effective.

The probabilistic sensitivity analysis incorporated the uncertainty around the postsurgical mortality observed in MASS; a one-way sensitivity analysis using the lower mortality rate observed in NAAASP, based on limited data, reduced the latter ICER by approximately £300. One-way sensitivity analysis suggests that the cost-effectiveness ratio would rise above the NICE £20 000 threshold at a prevalence of AAA in 65-year-old men of 0·35 per cent, compared with the observed 1·5 per cent.

Discussion

To assess the cost-effectiveness of many interventions, particularly screening where the bulk of costs are upfront, but benefits are accrued over time, long-term modelling is essential. It is rare to be able to revisit a model originally constructed using short-term (4-year) trial evidence and compare modelled results with more robust mid-term (10-year) trial data. Such models may not, however, as here, predict well over the medium term. The efforts to recalibrate the model confirmed that the cost-effectiveness estimates are more sensitive to the modelled differences between arms in costs and outcomes (incremental costs and QALYs) than the absolute values in each arm. For that reason, the focus of calibration should be on these differences that drive the cost-effectiveness ratio. The revalidation exercise undertaken demonstrates that economists should be cautious in the use of models based on relatively short-term data[13], given that they may not extrapolate well to medium- or long-term outcomes.

These new analyses have not simply been updated to reflect longer-term trial data. Data from recent meta-analyses of aneurysm rupture and growth rates were used to estimate the growth and rupture rates over the long term. New unit cost estimates for the screening procedure and for AAA surgery that reflect current practice in the UK were incorporated. The new cost estimates demonstrate that, although simple adjustment using relevant price indices may be adequate for some unit costs, for some the procedure costs need to be re-estimated to reflect changes in the costs of particular resources, and changes in the process of care.

Most importantly from a policy perspective, the model incorporates key parameters from the first years of NAAASP: attendance, AAA prevalence and size distribution at first screen. The combined changes do mean that the estimated 30-year ICER of £7370 per QALY gained has increased; the original model estimated an ICER of £2970 per QALY gained[8]. The increase in the estimated ICER reflects the incorporation into the modelling of the much lower AAA prevalence found by NAAASP (1·5 per cent) compared with MASS (4·9 per cent). It also reflects, as might be expected, the fact that the cost of screening has increased since the first costing exercise was conducted in 2001. The costs of elective and emergency AAA repair have increased well above general health service inflation, in part due to the use of more expensive EVAR procedures.

Despite the increase in the estimated ICER, the new modelling demonstrates with confidence that AAA screening remains highly cost-effective, with an ICER well below the lower limit of NICE's acceptable cost-effectiveness range of £20 000–30 000 per QALY gained. The probabilistic sensitivity analysis suggests that, even at a level of £10 000 per QALY, the probability that NAAASP is cost-effective is 99 per cent, thus providing strong support for cost-effectiveness of the current screening programme in the UK.

Although early estimates of the cost-effectiveness of AAA screening predating the publication of results from randomized trials were very variable[22], and precise estimates of cost-effectiveness are necessarily country-specific, there is now a growing international consensus that one-off ultrasound screening in men at around age 65 years is cost-effective. This conclusion for the UK is paralleled by studies relating to Canada[23], Denmark[24,25], The Netherlands[26], Norway[26], Northern Ireland[27] and Italy[28], with only one recent contrary estimate, also from Denmark[29].

Acknowledgements

The authors thank the RESCAN Collaboration[11] for providing estimates of small AAA growth and rupture rates, M. Bown and J. Powell for helpful discussions, the NHS AAA screening programme for additional unpublished data relating to NAAASP, and E. Diment for data from the National Vascular Database.

This project was supported in part by the UK National Institute for Health Research Health Technology Assessment Programme (project 08/30/02). The funders played no role in the conduct of the study, in the collection, management, analysis and interpretation of the data, or in the preparation and review of the manuscript. The views expressed in this article are those of the authors and not necessarily those of the UK National Health Service, UK NAAASP or the UK National Screening Committee. M.J.B. has been a member of the UK National Screening Committee since 2009.

Disclosure: The authors declare no conflict of interest.

References

1 Ashton HA, Buxton MJ, Day NE, Kim LG, Marteau TM, Scott RA *et al*. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *Lancet* 2002; **360**: 1531–1539.

2 Kim LG, P Scott RA, Ashton HA, Thompson SG; Multicentre Aneurysm Screening Study Group. A sustained mortality benefit from screening for abdominal aortic aneurysm. *Ann Intern Med* 2007; **146**: 699–706.

3 Thompson SG, Ashton HA, Gao L, Scott RA; Multicentre Aneurysm Screening Study Group. Screening men for abdominal aortic aneurysm: 10 year mortality and cost

effectiveness results from the randomised Multicentre Aneurysm Screening Study. *BMJ* 2009; **338**: b2307.
4 Thompson SG, Ashton HA, Gao L, Buxton MJ, Scott RA; Multicentre Aneurysm Screening Study Group. Final follow-up of the Multicentre Aneurysm Screening Study (MASS) randomized trial of abdominal aortic aneurysm screening. *Br J Surg* 2012; **99**; 1649–1656.
5 Cosford PA, Leng GC. Screening for abdominal aortic aneurysm. *Cochrane Database Syst Rev* 2007; (**2**)CD002945.
6 US Preventive Services Task Force. Screening for abdominal aortic aneurysm: recommendation statement. *Ann Intern Med* 2005; **142**: 198–202.
7 Multicentre Aneurysm Screening Study Group. Multicentre aneurysm screening study (MASS): cost effectiveness analysis of screening for abdominal aortic aneurysms based on four year results from randomised controlled trial. *BMJ* 2002; **325**: 1135.
8 Kim LG, Thompson SG, Briggs AH, Buxton MJ, Campbell HE. How cost-effective is screening for abdominal aortic aneurysms? *J Med Screen* 2007; **14**: 46–52.
9 UK National Screening Committee. *The UK NSC Policy on Abdominal Aortic Aneurysm Screening in Men Over 65*. http://www.screening.nhs.uk/aaa [accessed 20 October 2013].
10 NHS Abdominal Aortic Aneurysm Screening Programme. *Annual Report 2011–12*. http://www.aaa.screeening.nhs.uk/annualreport [accessed 11 August 2013].
11 RESCAN Collaboration, Bown MJ, Sweeting MJ, Brown LC, Powell JT, Thompson SG. Surveillance intervals for small abdominal aortic aneurysms: a meta-analysis. *JAMA* 2013; **309**: 806–813.
12 Sweeting MJ, Thompson SG, Brown LC, Powell JT; RESCAN Collaborators. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *Br J Surg* 2012; **99**: 655–565.
13 Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. *Health Econ* 2010; **19**: 43–55.
14 Thompson SG, Brown LC, Sweeting MJ, Bown MJ, Kim LG, Glover MJ *et al*. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. *Health Technol Assess* 2013; **17**: 1–118.
15 Brown LC, Powell JT, Thompson SG, Epstein DM, Sculpher MJ, Greenhalgh RM. The UK EndoVascular Aneurysm Repair (EVAR) trials: randomised trials of EVAR *versus* standard therapy. *Health Technol Assess* 2012; **16**: 1–218.
16 The Vascular Society of Great Britain and Ireland. *Abdominal Aortic Aneurysm Quality Improvement Programme*; 2012. http://www.aaaqip.com/aaaqip/data.html [accessed 4 July 2013].
17 Scott R, Vardulaki K, Walker N, Day N, Duffy S, Ashton H. The long-term benefits of a single scan for abdominal aortic aneurysm (AAA) at age 65. *Eur J Vasc Endovasc Surg* 2001; **21**: 535–540.
18 Office for National Statistics. *Interim Life Tables, England and Wales, 2009–2011*. http://www.ons.gov.uk/ons/rel/lifetables/interim-life-tables/2009-2011/stb-2009-2011.html [accessed 20 October 2013].
19 Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biom J* 2011; **53**: 750–763.
20 Kind P, Hardman G, Macran S. *UK Population Norms for EQ-5D. Discussion Paper 172*. Centre for Health Economics, University of York: York, 1999.
21 National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf [accessed 20 October 2013].
22 Campbell H, Briggs A, Buxton M, Kim L, Thompson S. The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm. *J Health Serv Res Policy* 2007; **12**: 11–17.
23 Montreuil B, Brophy J. Screening for abdominal aortic aneurysms in men: a Canadian perspective using Monte Carlo-based estimates. *Can J Surg* 2008; **51**: 23–34.
24 Lindholt JS, Sørensen J, Søgaard R, Henneberg EW. Long-term benefit and cost-effectiveness analysis of screening for abdominal aortic aneurysms from a randomized controlled trial. *Br J Surg* 2010; **97**: 826–834.
25 Søgaard R, Laustsen J, Lindholt J. Cost effectiveness of abdominal aortic aneurysm screening and rescreening in men in a modern context: evaluation of a hypothetical cohort using a decision analytical model. *BMJ* 2012; **345**: e4276.
26 Spronk S, van Kempen BJ, Boll AP, Jørgensen JJ, Hunink MG, Kristiansen IS. Cost-effectiveness of screening for abdominal aortic aneurysm in the Netherlands and Norway. *Br J Surg* 2011; **98**: 1546–1555.
27 Badger SA, Jones C, Murray A, Lau LL, Young IS. Implications of attendance patterns in Northern Ireland for abdominal aortic aneurysm screening. *Eur J Vasc Endovasc Surg* 2011; **42**: 434–439.
28 Giardina S, Pane B, Spinella G, Cafueri G, Corbo M, Brasseur P *et al*. An economic evaluation of an abdominal aortic aneurysm screening program in Italy. *J Vasc Surg* 2011; **54**: 938–946.
29 Ehlers L, Overvad K, Sørensen J, Christensen S, Bech M, Kjolby M. Analysis of cost effectiveness of screening Danish men aged 65 for abdominal aortic aneurysm. *BMJ* 2009; **338**: b2243.

APPENDIX B: Paper 2

Paper 2: AAA surveillance intervals

Thompson SG, Brown LC, Sweeting MJ, Bown MJ, Kim LG, Glover MJ, Buxton MJ, Powell JT and the RESCAN collaborators. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. Health Technol Assess. 2013;17(41):1-118. https://doi.org/10.3310/hta17410

Extract presented here is: 'Chapter 7: Methods for cost-effectiveness analysis of alternative surveillance policies' and 'Chapter 8: Results of cost-effectiveness analysis of alternative surveillance policies'

Chapter 7 Methods for cost-effectiveness analysis of alternative surveillance policies

In this chapter we describe the methods and preparatory work involved in assessing the cost-effectiveness of different surveillance policies. This includes how the information on small AAA growth rates and rupture rates from the IPD meta-analyses (see *Chapters 5* and *6*) is used to provide inputs for this health economic modelling. The results of the cost-effectiveness analyses are given in *Chapter 8*.

Introduction

Previous studies have estimated the cost-effectiveness of the screening strategy that was evaluated in the MASS trial. Estimates of the cost-effectiveness of this strategy have been adjusted over time, principally to reflect the emerging long-term data from the trial. The original cost-effectiveness estimates were based simply on 4 years of follow-up and estimated a mean cost per life-year gained within that truncated period of £28,400 (95% CI £15,000 to £146,000)[56] at 2000–1 prices. Using the observed data at 10 years of follow-up, the estimate had fallen to £7600 (95% CI £5100 to £13,000) despite revaluing costs to 2008–9 prices.[4] Based on informal modelling, the original cost-effectiveness paper had suggested that the cost-effectiveness over 10 years would indeed be around £8000 per life-year saved.[56]

However, it is clearly recognised that an investment in a screening programme needs to be assessed over a longer period that does not cut short the benefits from avoided aneurysm-related mortality and will require a formal model. Such a model, taking a 30-year perspective but initially based on the 4-year MASS results, was developed and this estimated the cost per life-year gained over a 30-year period as £2320 (with a 95% uncertainty interval of £1600 to £4240).[8] As the period of follow-up of the MASS trial has extended, the uncertainty associated with the longer-term effects of screening has been reduced, and scope now exists to compare the model results at 10 years and as necessary to recalibrate the model to more accurately reflect the longer-term observed data.

All these estimates have assumed a screening programme in elderly men with a surveillance pattern of: no recall for patients with an aortic diameter of < 3.0 cm; yearly rescanning of patients with an aortic diameter of 3.0–4.4 cm; 3-monthly rescans for patients with an aortic diameter of 4.5–5.4 cm; and consideration for surgery if an aortic diameter of ≥ 5.5 cm. This surveillance strategy adopted in MASS was based on the data available and expert clinical opinion at the time of the planning of the trial in 1995–6. The same pattern was subsequently adopted by the NAAASP. The analysis of growth and rupture rates in this RESCAN project provides an evidence base and opportunity to investigate whether or not a different surveillance strategy might be better than that used in MASS and subsequently by NAAASP.

Other economic models of AAA screening have of course been published. A systematic review of the models to 2006 emphasised the variability in their estimates of cost-effectiveness.[57] Since that review, a number of further modelling studies have been published drawing on data from a variety of sources: for example studies relating to Italy,[15] to the Netherlands and Norway,[58] to Canada[59,60] and to Denmark.[61] These have all concluded that screening is acceptably cost-effective, with the exception of a study using data from Denmark and other sources, which concluded that screening did not seem to be cost-effective.[62] A recent modelling study, again using Danish data, contradicts that previous conclusion, suggesting that screening men at age 65 years is highly cost-effective compared with no screening and, additionally, that rescreening after 5 years may be a cost-effective extension to the programme.[63] However, no models of cost-effectiveness of AAA screening have been published that have specific relevance to NAAASP, other than those already cited relating to and derived from the MASS study, which is the largest randomised trial of AAA screening and contributes most to the international evidence.[7]

The focus of the cost-effectiveness element of the current study is to analyse the implications of the RESCAN analyses, relating to the international evidence on growth rates and rupture rates, for the most cost-effective surveillance strategy following AAA screening. Cost-effectiveness is of course a function not only of these clinical parameters but also the behavioural, resource-use, and cost data relevant to a particular screening programme. Hence, we have chosen to use such country-specific data relevant to the current UK programme.

The work specifically undertaken on the model for this study consisted of the following elements:

1. Validation and recalibration of the previously published model[8] to best reflect the accumulated 10-year follow-up data from MASS, including incorporation of time-dependent parameter values.
2. Adaption of the model structure to fully incorporate unobserved 'tunnel' states to reflect the growth and rupture probabilities for all aneurysms, whether or not reobserved and remeasured at a recall scan.
3. Re-estimation of current unit costs for screening interventions and elective and emergency surgery.
4. Incorporation of data from NAAASP, including attendance rates, prevalence of identified aneurysms, and distribution of aneurysm sizes at initial screening.
5. Incorporation of size-specific growth and rupture rates from the analysis of the IPD surveillance data sets reported in *Chapters 5* and *6*.

Each of these aspects of our development of the model is to enable us to address the issue of the cost-effectiveness of alternative surveillance strategies, in the context of the NAAASP, and each aspect is addressed in turn in this chapter.

The initial model

The starting point for the economic analysis in this project was the economic model that had previously been developed by members of the research team, to estimate the long-term cost-effectiveness of a screening programme using the MASS trial screening and surveillance protocol. Details of this model have previously been published.[8] In summary, it used a Markov model, with 3-monthly cycles, to compare the introduction of a formal screening programme of an invitation for a one-off US scan for men aged 65 years, and the MASS surveillance strategy described above if an aneurysm is identified, with the policy of no systematic screening. The original model is represented in *Figure 18*.

The model used a 30-year time horizon (so approximating lifetime results). The methods were consistent with those recommended for the National Institute for Health and Care Excellence (NICE)'s Technology Appraisal Programme.[64] The cost perspective was that of the NHS and the model estimated life-years gained and adjusted these for age-specific utility values relating to the UK population to provide ICERs in terms of QALYs.

For this model, parameter values were estimated from patient-level data from the 4-year analysis of the MASS study and the model structure was internally validated by comparing its results against observed event and cost-effectiveness data over the 4-year time period. In addition, it was externally validated against key outcomes from the Cochrane review of AAA screening and surveillance.[7]

Validation and recalibration of the previously published model to better reflect the 10-year follow-up data from the Multi-centre Aneurysm Screening Study

The first step in the current study was to revalidate the model comparing the 10-year model outcomes against the 10-year observed data from the follow-up of MASS, and, where necessary, to adjust or 'recalibrate' the model parameters to better reflect that observed data. We used the 8.9–11.2 years of observed follow-up (mean 10.1 years) to better estimate the values for the parameterisation of the economic

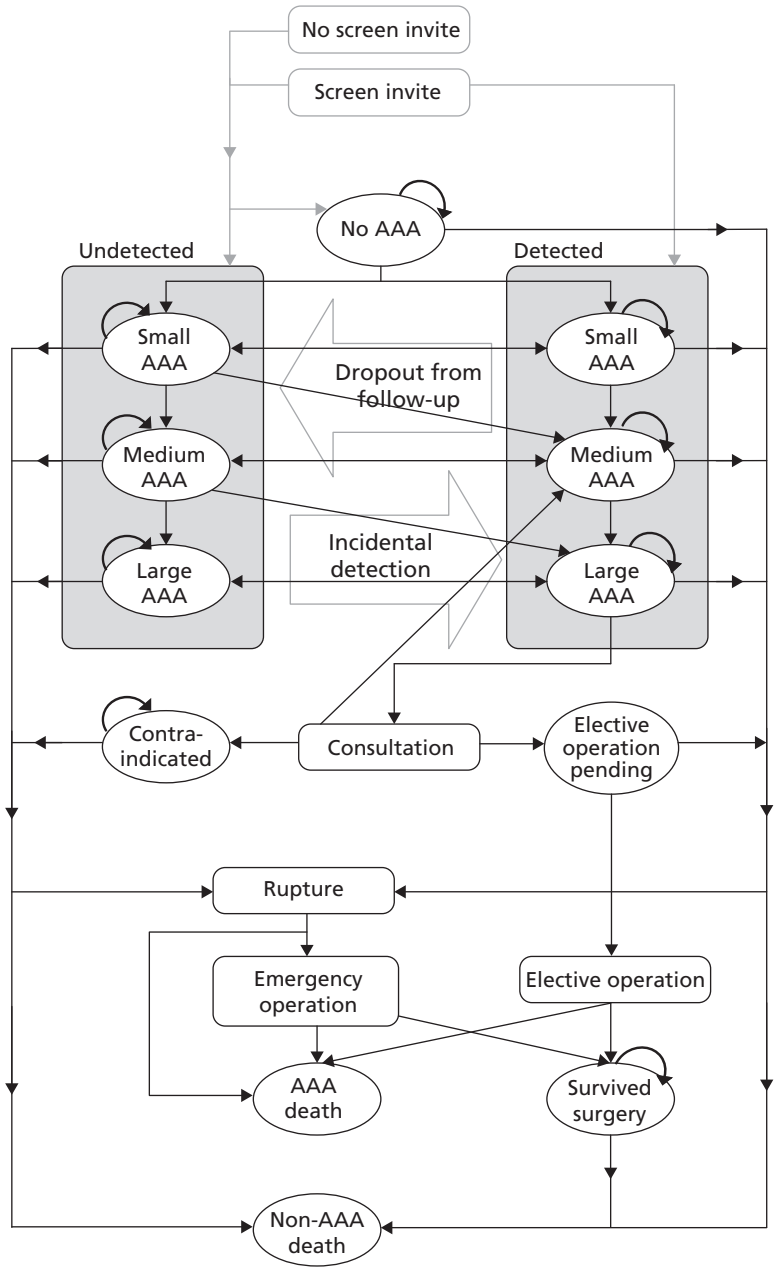


FIGURE 18 Original health economic model structure. Reproduced with permission from figure 1 of Kim LG, Thompson SG, Briggs AH, Buxton MJ, Campbell HE. How cost-effective is screening for abdominal aortic aneurysms? *J Med Screen* 2007;**14**:46–52. http://dx.doi.org/10.1258/096914107780154477.[8]

model. Outputs of the model were compared against those observed in the trial, by considering the numbers of key observed events and cost-effectiveness results at the end of this follow-up period.

Given the staggered recruitment of participants into the trial and the range of follow-up periods for individual patients, for the purpose of comparing numbers of key events the model replicated censoring patterns by removing equal proportions as observed in the trial of numbers in each state for every 3-month cycle, so that the total person-years of follow-up were comparable. The model was then run over 11.25 years to provide comparability. To produce estimates of costs and life-years at 10 years comparable with those observed in MASS, the model was run for 10 years with costs based on 2008–9 prices and survival based on all-cause mortality. Given the characteristics of the MASS trial population, it might be expected that mortality

rates would not be comparable with the national statistics, and so for the internal validation exercise MASS-specific mortality data were used to estimate the probability of mortality in each 3-month cycle, rather than using national mortality statistics that were incorporated into the model to give it greater external validity or generalisability.

Initially the model did not appear to perform particularly well against observed data in terms of key events or cost-effectiveness results over the longer follow-up period. As the length of follow-up had increased, trends may have emerged in key parameter values and differences between modelling and observed outputs could be partly attributed to the use of time-constant parameters. Time-dependent transition probabilities were therefore estimated using 'logistic' and 'Poisson' regressions, to improve the fit of the model. The differences between the observed and modelled data may also have been in part due to parameters used to inform the modelling that could not be observed in MASS. In particular, two parameters, the probability of opportunistic detection among those not in active screening and the rupture rate of undetected large aneurysms, were seen as potentially unreliable. The first had been based on calculations utilising data from the control arm of the trial, to give a crude estimate of the detection rate. The latter had been estimated to fit the 4-year data, assuming it lay between the rupture rate in detected large aneurysms and the rupture rate among aneurysms contraindicated for elective surgery. A recalibration exercise was conducted using a range of figures for both these parameters, to obtain model results that gave more similar numbers of key events (i.e. elective operations, emergency operations and AAA deaths) to those observed in MASS.

It was not easy to achieve complete consistency between the observed and modelled clinical outcomes. Given that the long-term ICER was the primary outcome for the cost-effectiveness modelling, the recalibration focused on eliminating disparities in the modelled and observed differences between the arms.

Table 13 shows key event rates as observed in MASS, as estimated by the original model and as estimated by the model following recalibration. The pattern of events over time from the observed data, the initial model estimates and the recalibrated estimates are shown in *Figure 19a–h*.

This recalibration process achieved similarity between the modelled and observed differences in key events between the arms, as well as in the resultant ICERs (*Table 14*). The ICER was £7600 per life-year based on observed 10-year data: the original model estimated an ICER at 10 years of £18,000, the recalibrated model estimated the ICER at £8900, closer to that from the observed data. This suggested that, though imperfect, the model would be suitable for extrapolation of cost-effectiveness results over the long term.

Adapting the model structure to fully incorporate unobserved 'tunnel' states

The previously constructed Markov model[8] needed adaption to account more explicitly for the incidence of rescans. The original model, with 3-monthly cycles, was built with a view to considering one surveillance policy (that pertaining to the MASS data of yearly recall for AAAs measuring 3.0–4.4 cm and 3-monthly recall for medium AAAs measuring 4.5–5.4 cm). The model averaged out the surveillance for the smaller aneurysms, effectively assuming that surveillance scans were being conducted every cycle but assigning only one-quarter of the rescan costs to each cycle. This meant that individuals were able to transition to larger aneurysmal states in each 3-month cycle and on reaching the large state they could be considered for elective surgery or returned to screening. However, in reality, those in a small AAA state (3.0–4.4 cm) are not rescanned every cycle and cannot move into the surveillance pattern for a larger aneurysm until the increased size of their aneurysm has been identified through a scan. In the original model, costs associated with rescanning were averaged across the cycles, 0.25 of a rescan cost per cycle for the small aneurysmal state and assigned as one per cycle for the medium-sized AAA. Although this approximation was adequate when considering the one surveillance strategy, the model needed to be adapted for the purposes of answering questions around different surveillance policies.

TABLE 13 Comparison of key events observed in MASS and economic model

10-year cumulative key events	Observed in MASS[a]	Original model[b]	Model after recalibration[c]
Control group			
Elective operations	226	256	213
Emergency operations	141	140	168
AAA deaths	296	305	385
Non-AAA deaths	10,185	10,139	10,148
Invited group			
Elective operations	552	607	539
Emergency operations	62	88	97
AAA deaths	155	202	248
Non-AAA deaths	10,119	10,185	10,189
Difference between invited and control groups			
Elective operations	326	351	326
Emergency operations	–79	–52	–71
AAA deaths	–141	–103	–137
Non-AAA deaths	–66	46	41

a Key events observed in MASS at 10-year follow-up.
b Economic model using time-constant parameter estimates from MASS 10-year follow-up.
c Economic model, with time-dependent parameter estimates from MASS 10-year follow-up and post-calibration exercise.

This adaption was achieved through the inclusion of 'tunnel' states for aneurysm growth, accounting for patients in whom aneurysm growth occurred but was not observed or acted on. The structure of the extended model is represented diagrammatically in *Figure 20*. These tunnel states allow aneurysms to grow in between scans, where the individual is subject to the relevant rupture rate. An individual can stay in the tunnel state that they have entered in subsequent cycles, or transition to larger AAA tunnel states. Only when a scheduled surveillance scan occurs are individuals then able to be move out of these unobserved tunnel states into observed states. On this they can then be subject to more frequent scanning, or events that occur as a result of entering the large AAA state (i.e. consideration for elective surgery). This provides a more accurate representation of reality and most importantly can allow for the effects of differing policies by adjusting the number of cycles that individuals spend in these states before a rescan and a return to observed states. Aneurysms that are detected opportunistically are assumed to enter an observed state before moving into tunnel states. However, given that this opportunistic detection can occur in every cycle, an approximation in the cycles immediately following opportunistic detection is used to bring these individuals into the same rescanning schedule as the rest of the detected aneurysm cohort. This retains programming efficiency and has the effect that some opportunistically detected patients will be assumed to receive their first 1-year scans after a shorter period than a full year.

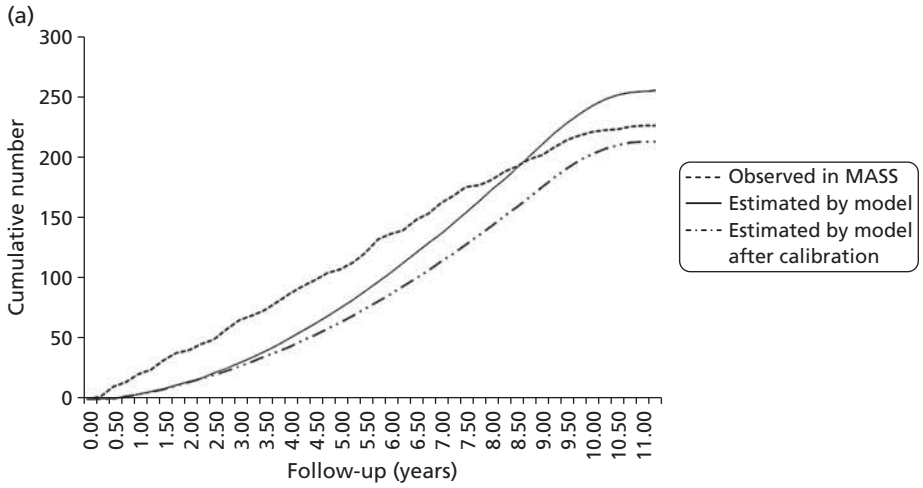


FIGURE 19a Number of elective operations in control group over 10 years' (mean) follow-up.

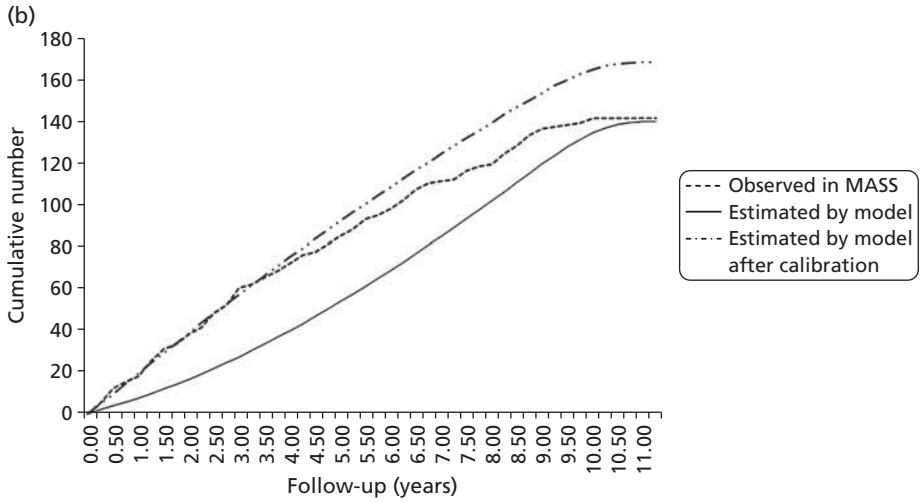


FIGURE 19b Number of emergency operations in control group over 10 years' (mean) follow-up.

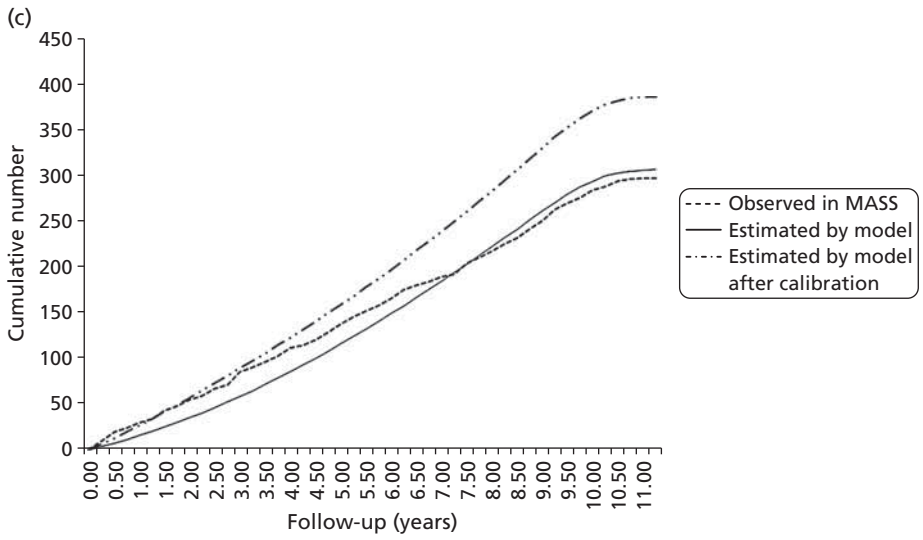


FIGURE 19c Number of AAA-related deaths in control group over 10 years' (mean) follow-up.

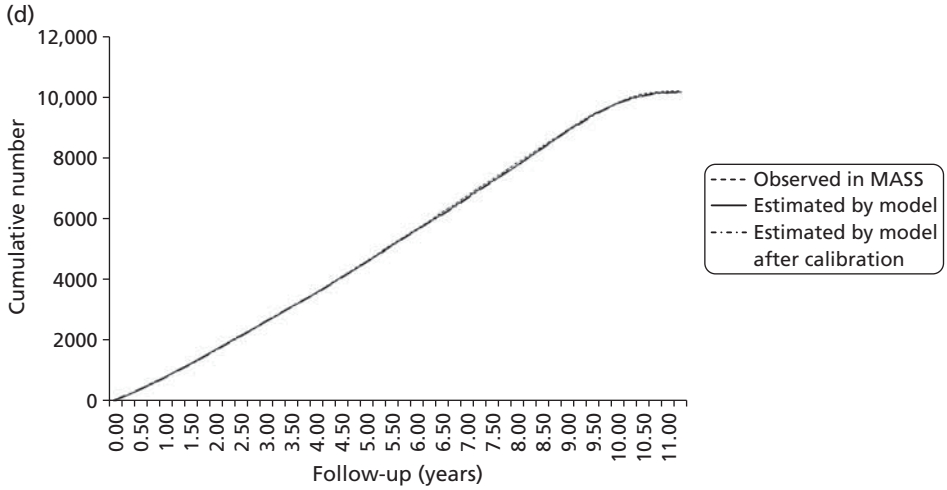


FIGURE 19d Number of non-AAA deaths in control group over 10 years' (mean) follow-up.

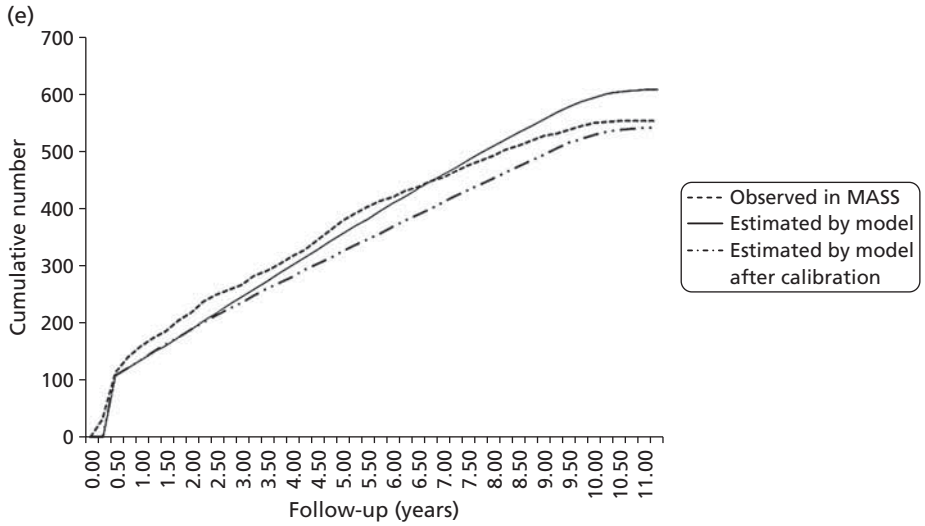


FIGURE 19e Number of elective operations in invited group over 10 years' (mean) follow-up.

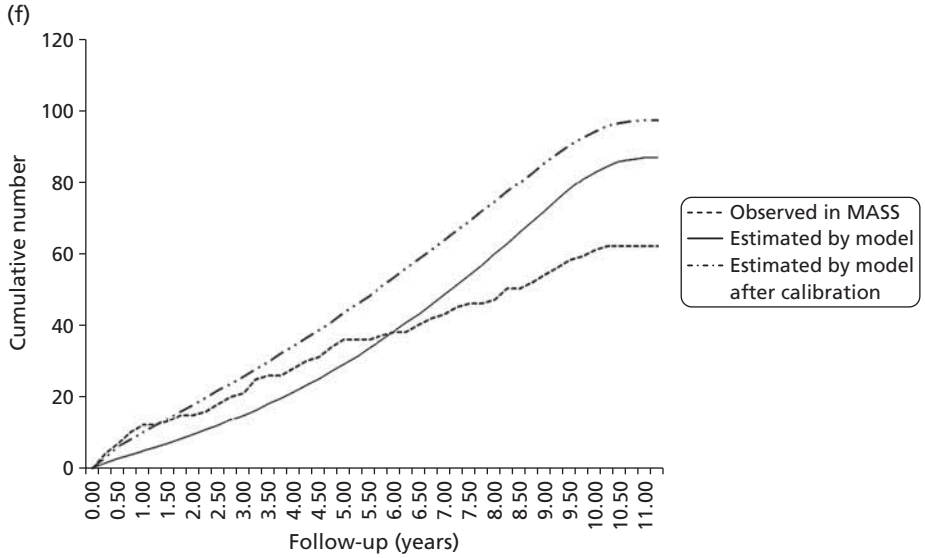


FIGURE 19f Number of emergency operations in invited group over 10 years' (mean) follow-up.

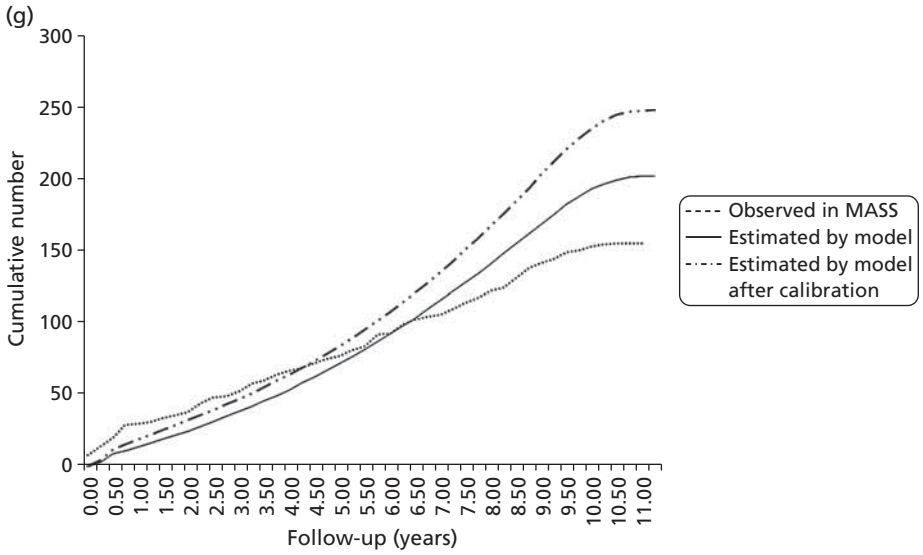


FIGURE 19g Number of AAA-related deaths in invited group over 10 years' (mean) follow-up.

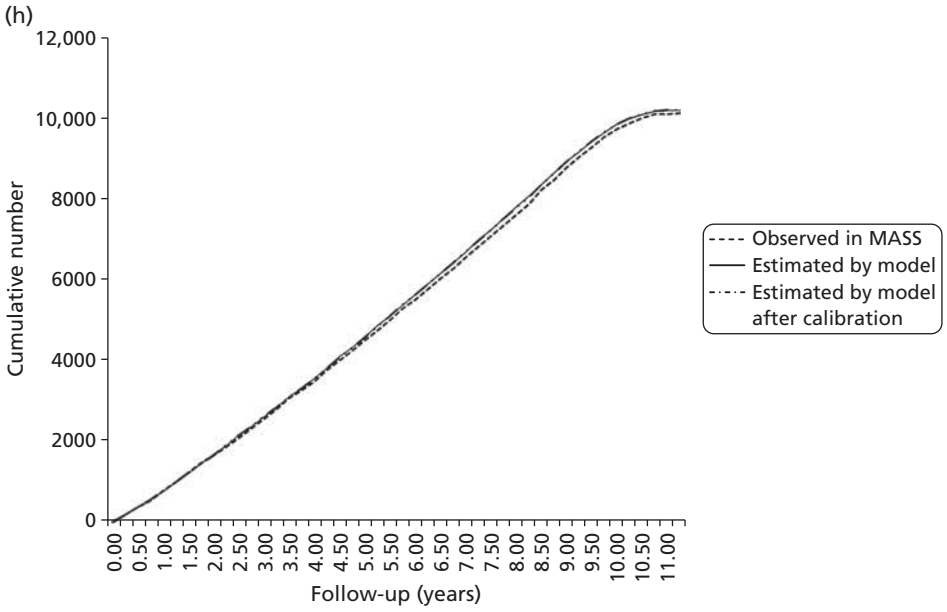


FIGURE 19h Number of non-AAA deaths in control group over 10 years' (mean) follow-up.

It may be useful to consider the current screening policy of 1 year and 3 months for small and medium AAAs, respectively, as an example of the way the new structure works. An individual invited to and having attended screening, with a small aneurysm observed at baseline, would be invited to return for another scan in 1 year, equivalent to four cycles in the model. Assuming the AAA had not grown, nor ruptured and the individual had not dropped out of surveillance or died, they would move into the small (unobserved) tunnel state in cycle 2. If no transition had occurred in cycles 3 and 4, the individual would then move out of the small unobserved 'tunnel' state back into the small observed state where a rescan takes place. They would then move into tunnel states again following this scan. Alternatively, the aneurysm could have grown and the patient transitions into the tunnel state for medium aneurysms in cycle 2 and possibly into the tunnel state for large aneurysms in cycle 3. The tunnel states mean that this individual will not be treated as having a large aneurysm as regards the surveillance policy until this growth has been observed at the next scan in cycle 4.

TABLE 14 Comparison of discounted mean costs and effects observed in MASS and estimated in economic model

Cost-effectiveness at 10 years[a]			
	Observed in MASS	Original model	Model after recalibration
Control group			
Life-years (mean)	7.509	7.291	7.282
Cost, £ (mean)	108	118	124
Invited group			
Life-years (mean)	7.523	7.297	7.293
Cost, £ (mean)	208	233	225
Difference between invited and control groups			
Difference in life-years, £ (mean)	0.014	0.006	0.011
Difference in costs, £ (mean)	100	115	101
ICER (life-years), £	7600	18,000	8900
ICER (QALYs), £[b]	9700	23,000	11,400

a Costs based on 2008–9 prices; costs and mortality discounted at 3.5%. Survival based on all-cause mortality.
b Life-years adjusted using population norms.

These techniques for modelling growth also mean that the accrual of costs due to rescanning should be more accurate. Costs will accrue as rescans occur over time, as opposed to averaging costs over a number of cycles. By implementing these changes the model can be used to investigate a range of potential new surveillance policies and compare their cost-effectiveness.

Re-estimation of current unit costs for screening interventions, and elective and emergency surgery

The various estimates of cost-effectiveness based on MASS, referred to earlier, have so far all used original unit cost estimates at 2000–1 prices, updated as necessary only for general health service inflation. There was a clear need for a more thorough update of costs to reflect a range of changes in practice. All costs were re-estimated at 2010–11 price levels.

Screening intervention costs

The original costs for the elements of screening came directly from the costs of the services in the trial. These have now been superseded by costs directly provided by the NAAASP. *Table 15* shows the original figures, those figures updated to 2010–11 price levels, and the current unit costs, again at 2010–11 price levels from the NAAASP.

Costs for elective and emergency aneurysm repair

The MASS estimates for the cost of elective and emergency procedures were calculated using very detailed bottom-up costing using patient notes and other detailed hospital records of 577 patients from the four surgery centres involved in the MASS study.[56] It was not feasible to re-estimate the costs using the same resource-intensive methods for this study. Rather we chose to estimate unit costs for surgery using more recent published studies where available and where more recent, relevant studies were not available by making explicit adjustments to the mean per patient costs previously calculated to reflect more recent routine data on resource use.

The most recent and largest (sample size) costing exercise was completed as part of the analysis of the endovascular aneurysm repair (EVAR)-1 trial, which we have used to estimate the costs of open repair and

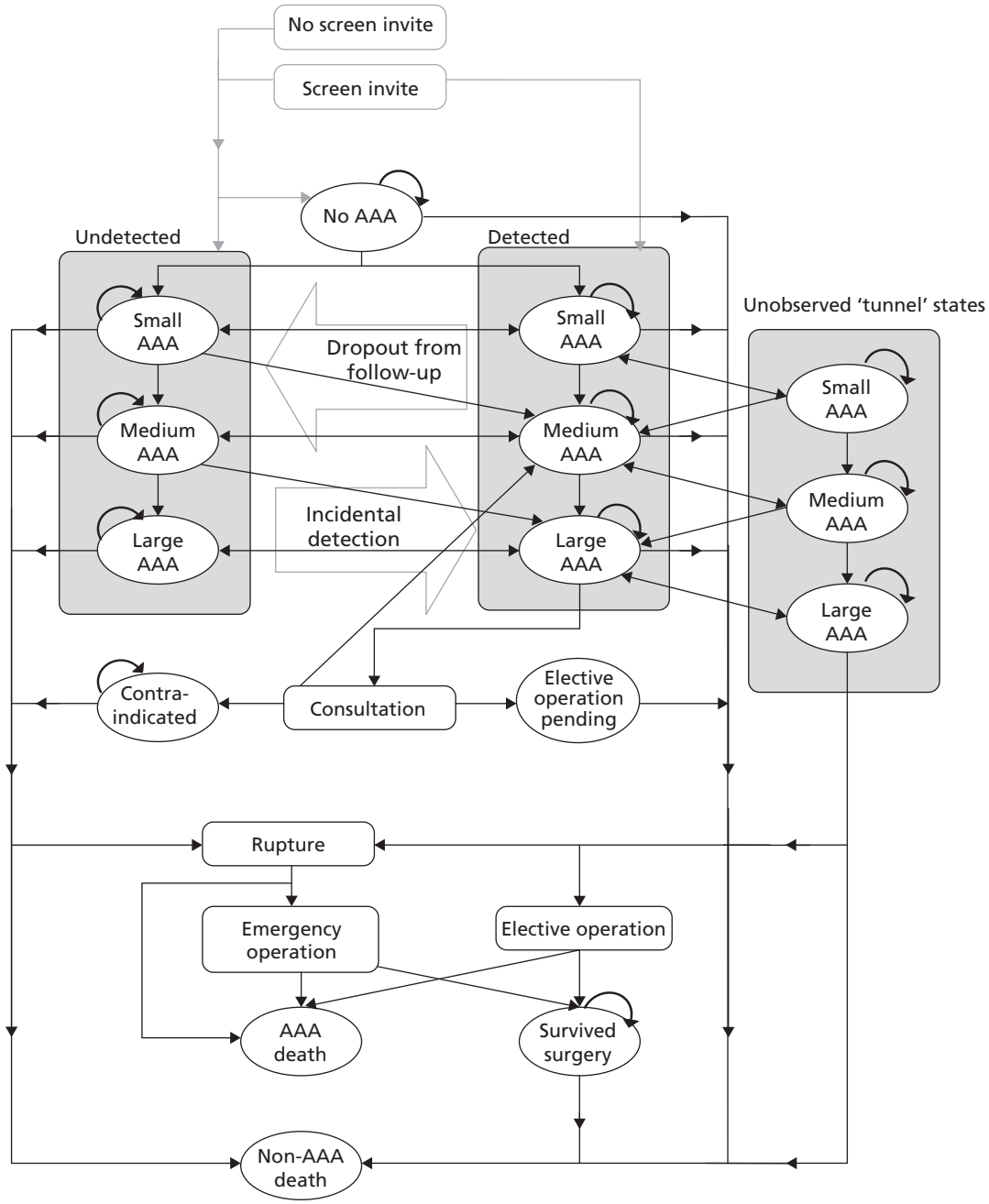


FIGURE 20 Extension of original model structure to include unobserved 'tunnel' states.

endovascular repair in an elective setting.[65] As per NICE guidance,[66] we assumed that emergency EVAR operations are only being performed in a research environment and, so, we have only included costs for open emergency repair in our estimates. For emergency open repair an estimate of per patient costs was updated from the MASS trial, in the following way:

(a) First, the major costs components were removed from the total mean per patient estimates for elective open repair and elective EVAR (from EVAR 1 trial[65] estimates) and emergency OR (from MASS trial[56] cost estimates). In each case, this constituted the removal of costs attributable to hospital stay and operation time. For EVAR this also included the cost of the stent. This cost was updated using more recent data from centres involved in the screening programme and correspondence with manufacturers.

TABLE 15 Costs for elements of screening

Element	Original MASS[56] cost 2000–1 (£)	MASS[56] cost inflated to 2010–11 (£)	Updated unit cost[a] (£)
Invitation to screen	1.31	1.84	1.70
Cost of first scan	19.08	26.80	32.20
Surveillance scan	46.04	64.67	68.00

a Source: UK NAAASP, personal communication, 2012.

(b) The residual cost, made up of more minor resource-use items, such as blood products used and consumables, was inflated to 2010–11 prices according to Personal Social Services Research Unit indices.[67]
(c) The resource-use data were then updated for operation length and total hospital stay (apportioned using the same proportions for intensive care unit (ICU) and normal ward as had been observed in EVAR 1 trial or MASS, respectively) using recent data provided for us from the National Vascular Database (January 9 March 2012).[68] New unit costs for ICU, high-dependency unit (HDU) and vascular ward stay were all obtained from NHS reference costs for 2010–11.[69] A new unit cost estimate per hour of time in operating theatre was obtained.[70] These updated estimates were applied to the new resource-use estimates.
(d) The elective cost has been weighted according to the proportion of cases that are EVAR and OR in the National Vascular Database sample (approximately 70/30 for EVAR and OR, respectively).

Tables 16–18 illustrate this process for updated components of resource use and unit costs, for elective open, elective EVAR, and for emergency procedures, respectively. *Tables 19–21* show the magnitude of the residual component of each of these three costs, which was simply inflated to 2010–11 prices. *Table 22* summarises

TABLE 16 Components of resource use and unit costs for elective open repair

Component	EVAR 1[65] resource use	EVAR 1[65] unit cost (£)	Updated resource use	Updated unit cost (£)
Theatre time	215 minutes	17.58	181.97 minutes	20.67
HDU	1.88 days	832.00	1.65 days	883.00
ITU	2.47 days	1165.00	2.16 days	1226.00
Vascular ward	11.41 days	268.00	9.98 days	266.00
Total days	15.76 days		13.79 days	

TABLE 17 Components of resource use and unit costs for elective EVAR

Component	EVAR 1[65] resource use	EVAR 1[65] unit cost (£)	Updated resource use	Updated unit cost (£)
Theatre time	191 minutes	17.58	147.04 minutes	20.67
HDU	0.83 days	832.00	0.63 days	883.00
ITU	0.59 days	1165.00	0.44 days	1226.00
Vascular ward	8.34 days	268.00	6.29 days	266.00
Total days	9.76 days		7.36 days	
Stent cost		5219.00		6500.00

TABLE 18 Components of resource use and unit costs for emergency open repair

Component	MASS trial[56] resource use	Updated resource use	Updated unit cost (£)
Theatre time	182 minutes	161.94 minutes	20.67
ITU	4.74 days	7.12 days	1226.00
Vascular ward	7.66 days	11.51 days	266.00
Total days	12.4 days	18.63 days	

TABLE 19 Calculation of residual costs component inflated to 2010–11 prices for elective open repair

Component	Residual cost (£)
Total open repair cost as per EVAR 1 trial[65]	11,842
Length of stay component	7214
Operation time component	3647
Total of components	10,861
Residual costs to inflate	981
Residual costs inflated to 2010–11 prices	1014

TABLE 20 Calculation of residual costs component inflated to 2010–11 prices for elective EVAR

Component	Residual cost (£)
Total EVAR cost as per EVAR 1[65]	13,019
EVAR stent and parts component	5219
Length of stay component	3543
Operation time component	3255
Total of components	12,017
Residual costs to inflate	1002
Residual cost inflated to 2010–11 prices	1036

TABLE 21 Calculation of residual costs component inflated to 2010–11 prices for emergency open repair

Component	Residual cost (£)
Total open repair cost as per MASS[56]	11,176
Length of stay component	6932
Operation time component	794
Total of components	7726
Residual costs to inflate	3450
Residual cost inflated to 2010–11 prices	4846

TABLE 22 Summary of updated unit costs for surgical procedures

Type of surgery	Updated cost per patient (£)	Cost per patient (EVAR 1[65]) (£)	Cost per patient inflated to 2010–11 prices (EVAR 1[65]) (£)	Cost per patient (MASS[56]) (£)	Cost per patient inflated to 2010–11 prices (MASS[56]) (£)
Elective OR	11,532.69	11,842.00	12,241.17	6909.00	9704.24
Elective EVAR	13,345.66	13,019.00	13,457.84		
Elective weighted	12,806.21				
Emergency OR	19,984.75			11,176.00	15,697.59

the new cost estimates and compares them with the original estimates, and the original estimates inflated to 2010–11 prices.

Costs of pre-surgical consultations

No newer estimates were available for the costs of consultations when referred to surgery which included costs of associated tests. The original MASS[56] estimates for these costs were simply inflated from 2000–1 to 2010–11 prices (£309.88 inflated to £435.25).

Incorporating data from the NHS Abdominal Aortic Aneurysm Screening Programme

To reflect current AAA screening in the UK, data from the NAAASP were utilised in the economic modelling. The attendance rate from those invited to screening in the NAAASP has been lower than that of the MASS trial (which was 83%); a figure of 73% from the NAAASP has been used in the modelling.[71] A lower prevalence of AAAs has also been noted in data from the NAAASP, which might have some effect on cost-effectiveness results. In the MASS trial the prevalence was 4.9%, but the current prevalence observed in the NAAASP is 1.6% and this rate was used in the modelling. Data were provided to us from the NAAASP on the distribution of aneurysm sizes detected in the screening programme, to reflect possible changes in aneurysm sizes observed at baseline compared with the MASS data used in the original model.

In addition, non-AAA mortality rates were estimated using data from Hospital Episodes Statistics[72] and the Office for National Statistics (ONS) for 2010,[73] incorporated into the modelling as age-specific 3-month probabilities.

Incorporating size-specific growth and rupture rates from the RESCAN analyses

Growth rates

From the reanalysis of existing surveillance data (see *Chapter 6*) we have developed random-effects models to describe AAA growth in each of the studies. These models are utilised to calculate how individuals pass through the size states of the Markov model used for the cost-effectiveness modelling. The methods are described fully in *Appendix 4*, with a summary given here.

Firstly, 3-month transition probabilities were calculated between 5-mm-wide size states (3.0–3.4 cm, 3.5–3.9 cm, 4.0–4.4 cm, 4.5–4.9 cm, 5.0–5.4 cm and 5.5+ cm). To achieve this, a cohort is envisaged with screening distribution of small aneurysm sizes taken from the Chichester screening study (shown to have an almost identical distribution to the NAAASP small diameters, as shown in *Figure 21*). This distribution is skewed towards small aneurysms with the mode close to 3.0 cm. The aneurysms in this envisaged cohort are

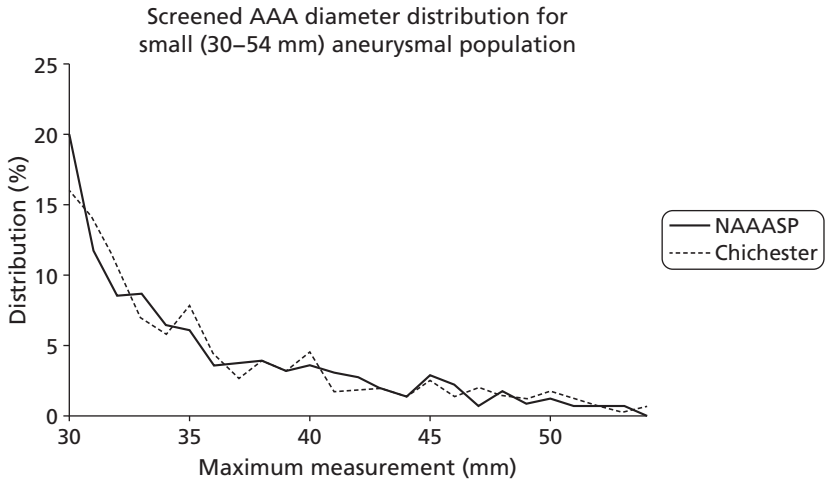


FIGURE 21 Distribution of AAA diameter at first screen from NAAASP data and Chichester Screening Programme data.

then allowed to grow over time using the baseline size-specific growth rates estimated from each study separately. For any 3-month period following screening, the estimated occupancy probabilities in each 5-mm-wide size state were calculated using the predictive distribution of aneurysm sizes. Those screened as normal are assumed to account for 98.3% of the screened population[9] and are also allowed to grow, with an estimated 0.207% reaching 3.0 cm every 3 months and, thereafter, growing at a rate equivalent to a screen detected patient with baseline diameter 3.0 cm. The occupancy probabilities from the cohort are then used to calculate 3-month transition probabilities between size states using the Chapman–Kolmogorov equations and assuming a progressive only Markov model.[74] The log-odds of the transition probabilities for each 3 months over a 30-year time period after screening are obtained together with their SEs, and these study-specific estimates are then pooled in a second stage using random-effects meta-analysis.

'Equivalence' probabilities

Different choices of small aneurysm states are considered in the Markov transition models to allow comparisons of different screening strategies. It is therefore necessary to calculate 3-month transition probabilities between any size states of interest. Given the 3-month transition probabilities calculated above for the 5-mm-wide size states we can obtain 'equivalent' transition probabilities for any concatenation of these states (e.g. 3.0–4.4 cm, 4.5–5.4 cm, 5.5+ cm). These 'equivalence' probabilities are defined by the requirement that two Markov models with different small aneurysm size states produce the same proportion of large aneurysms (> 5.5 cm) over time if no external intervention, deaths or censoring takes place. Hence, this facilitates a fair comparison of the different surveillance strategies since the rate of growth to large AAAs without intervention will be the same across all models. The method used to calculate these 'equivalence' probabilities is described in *Appendix 4*.

Rupture rates

To obtain 3-month transition probabilities of rupture over a 30-year time period, as required by the health economic model, we fit a parametric survival model to the data in each study. Only men are considered for this analysis (since men are the focus of the screening policy). Specifically, a Weibull proportional hazards model with time-updated covariate (AAA diameter) is used. Events other than rupture that terminate follow-up (lost to follow-up, non-rupture-related death or surgery) are classified as censored observations. The (transition) probability of rupturing over any 3-month time period given AAA diameter at the beginning of the period is then approximated from the survival distribution. The log-odds of these probabilities are obtained together with their SEs, and these study-specific estimates are pooled in a second stage using random-effects meta-analysis.

Analysis of cost-effectiveness using the adapted and updated model

These developments enable us to use the model to estimate the life-years, QALYs, costs and net monetary benefits for a series of strategies with different recall frequencies, compared initially with the base case of the existing strategy, and by looking at differences in net benefit to extend to comparisons between any two strategies.

In this particular context, a probabilistic sensitivity analysis (PSA), as might typically be provided as a representation of the overall uncertainty in the estimates of cost-effectiveness, would not adequately characterise that overall uncertainty. In particular, it is not readily feasible to estimate the correlated uncertainty around the very many (480) age- and size-related aneurysm growth and rupture rates derived from the IPD meta-analysis incorporated into the model, or to provide evidence-based estimates of uncertainty distributions around other parameters, or to characterise the underlying structural uncertainty in this complex model. Therefore, rather than provide a potentially misleading PSA, we have chosen to employ simple one-way sensitivity analyses to characterise the effects on net benefit of uncertainty around growth rate and rupture rate estimates, and to check whether or not the conclusions regarding the preferred strategy are sensitive to this uncertainty. For comparison we illustrate the effects of uncertainty around other important parameters in the model.

Chapter 8 Results of cost-effectiveness analysis of alternative surveillance policies

The current surveillance strategy used in the NAAASP follows that of the MASS trial. The strategy was based on the expert judgement informed by the limited data available at the planning stage of the trial. The analysis in this study of growth rates and rupture rates enable us now to analyse the cost-effectiveness of alternative surveillance in terms of the frequency of recall for men screened as having an aneurysm between 3.0 and 5.4 cm.

In this chapter we present the comparative cost-effectiveness of a number of alternative surveillance strategies using the revalidated, and extended model with updated unit costs and key characteristics from the NAAASP (uptake rates and distribution of aneurysm sizes at screening).

Alternative surveillance strategies

We used the adapted and updated model described in *Chapter 7* to examine the health benefits (in terms of life-years) and overall costs, modelled over a 30-year period, for a range of alternative surveillance strategies as compared with the current strategy which we refer to as strategy A. *Table 23* summarises the values and sources of clinical and cost parameters used. The alternative strategies included both lengthening and shortening the current time intervals between rescans for those identified with aneurysms between 3.0 and 4.4 cm (base case 1 year) and 4.5 and 5.4 cm (base case 3 months). We assumed that a recall more frequent than 3 months would be unworkable, so did not consider any such options. In analysing any of the different surveillance strategies we assume that the same strategy would also apply to all opportunistically identified aneurysms in both arms. This assumption explains the very small difference in the control arm life-years and cost between strategies.

Table 24 sets out the key comparison, which is illustrated on the cost-effectiveness plane in *Figure 22*. Strategy A represents the base case of the current screening strategy, with an ICER against no formal screening of £5572 per life-year gained and £7143 per QALY gained. To identify whether or not an alternative strategy was an improvement, and which was best, we focused on the ICERs (for both life-years and QALYS) and the net monetary benefit calculated as (Net QALYs × £20,000)–Net costs. We used the threshold value of £20,000 per QALY to be consistent with the opportunity cost of interventions within the NHS as articulated by NICE.[75]

Each of the alternative strategies demonstrated appropriate directional changes in the life-years gained from screening and the QALYs gained from screening. The strategies, which extend one or more intervals, slightly reduce the QALY gain (by missing a small proportion of aneurysms that would go on to rupture) and slightly reduce the cost (by avoiding additional scans), whereas strategies that decrease intervals increase the QALY gain and increase the costs. This effect is shown in *Figure 22*, where the strategies that increase intervals (strategies B, C, D and G) fall in the south-west quadrant of the cost-effectiveness plane (where both costs and QALYs are reduced), whereas a strategy that decrease intervals (strategy E) falls in the north-east quadrant (where both QALYs and costs increase). Strategy F, in which all aneurysms between 3.0 and 5.5 cm are recalled at 6-monthly intervals, can be eliminated from further consideration in that it reduces QALYs and increases costs, so is dominated by existing strategy A which is clinically more effective and cheaper.

The economic question for strategy E is whether or not the increase in QALYs is sufficient to justify the increase in costs, while for each of strategies B, C, D and G it is whether or not the reductions in cost are sufficient to compensate for the QALY losses. This is assessed against a view of the acceptable cost per QALY or threshold. Using the threshold of £20,000 per QALY, we can see that strategy E far exceeds the acceptable threshold and is not cost-effective. Given the only other practical strategy, to increase frequency

TABLE 23 Parameter estimates for adapted economic model

Clinical parameters	Estimate	Source
Proportion reinvited to screening	0.1360	MASS[8]
AAAs at first screen – attenders	0.0166	NAAASP[71]
AAAs at first screen – non-attenders	0.0166	NAAASP[71]
Non visualised AAAs	0.0166	NAAASP[71]
Proportion of scans non-visualised	0.0121	MASS[8]
Proportion of screen invited attending	0.730	NAAASP[71]
Proportion of AAAs at first screen – small	0.809	NAAASP[71]
Proportion of AAAs at first screen – medium	0.106	NAAASP[71]
Proportion of AAAs at first screen – large	0.0854	NAAASP[71]
Transition probabilities (3-monthly)		
Grow from no AAA to small AAA	0.00207	Chichester[53]
Grow from small AAA to medium AAA	TDTP[a]	RESCAN
Grow from medium AAA to large AAA	TDTP	RESCAN
Probability of dropout	0.0142	MASS
Rupture probability – no AAA	0	Assumption
Rupture probability – small AAA	TDTP	RESCAN
Rupture probability – medium AAA	TDTP	RESCAN
Rupture probability – detected large AAA	0.0125	MASS
Rupture probability – undetected large AAA	0.0282	Calibration
Rupture probability – contraindicated for surgery	0.0282	MASS
Probability of opportunistic detection	0.0114	Calibration
Probability of emergency surgery following rupture	0.368	MASS
Probability of death following emergency surgery	0.342	MASS
Proportion of large AAAs receiving surgery	0.681	MASS
Proportion of large AAAs returned to screening	0.221	MASS
Proportion of large AAAs contraindicated for elective surgery	0.0977	MASS
Probability of death following elective surgery – screen detected	0.0298	MASS
Probability of death following elective surgery – opportunistically detected	0.0717	MASS
All-cause mortality – contraindicated for surgery	0.0599	MASS
Age-specific all-cause mortality	Age specific	ONS[73]
Cost parameters		
Invitation to screen (£)	1.70	See *Chapter 7*
Cost of first scan (£)	32.20	See *Chapter 7*
Surveillance scan (£)	68.00	See *Chapter 7*
Pre-surgical consultation (£)	435.25	See *Chapter 7*
Elective repair (£)	12,806.21	See *Chapter 7*
Emergency repair (£)	19,984.75	See *Chapter 7*

TDTP, time-dependent transition probabilities.
a Time-dependent transition probabilities, applied to each 3-month cycle.

of scans, strategy F, was dominated by the present screening strategy, we can conclude that decreasing recall intervals is unlikely to be cost-effective. Looking at the strategies (B, C, D and G) that increase screening intervals, we are seeking a strategy that releases at least £20,000 in cost savings per QALY lost. All options meet that test, and of these we need to choose the one with the highest net benefit (highest excess in value of cost savings minus value of lost QALYs). Strategy C, with a recall pattern of 2 years for aneurysms between 3.0 and 4.4 cm and 3 months for aneurysms between 4.5 and 5.4 cm, has the highest net benefit, and is the most cost-effective strategy. Increasing intervals beyond this does not provide sufficient additional cost savings to justify the additional QALY losses.

To summarise these figures as clearly as possible, our analysis suggests that the effect of changing from the current surveillance strategy to surveillance strategy C, the best of the options considered, would mean that the loss of value from the QALY gain per man invited would be equivalent to a loss of £1.24, but that the cost would be reduced by £2.57 per man invited, giving a net monetary benefit gain of £1.33.

Uncertainty and sensitivity

As was indicated in *Chapter 7*, in this particular case it is not feasible, given available data and analytical possibilities, to provide a robust, meaningful PSA that encompasses all the parameter and structural uncertainty. We have therefore focused on a more appropriate series of one-way sensitivity analyses. These are summarised in *Table 25*.

The first of these applies growth and rupture rates drawn from the three UK population-based screening studies that used internal aortic diameter measurements (as does the NAAASP), namely MASS, Gloucester and Chichester (see *Chapter 6*), and which could therefore be seen as most applicable to the UK programme. Strategy C clearly remains the preferred option and, although it is slightly less cost-effective than in the base case, the net benefit is still positive and substantial. As an alternative way to address the uncertainty around growth and rupture rates, the next sensitivity analyses reduce or increase growth rates by 10% and rupture rates by 30%. These ranges approximate one SD in the estimates of the value of the relevant parameter. Again, in each case, the preference for strategy C, over all other strategies, remains unchanged. Using an estimate of operative mortality rates from the NAAASP, similarly does not change the choice of strategy. Further sensitivity analyses consider the possibility that dropout rates (from recall) might be affected by different recall strategies but again the preference for strategy C is unaltered. Finally, we consider the effect of (arbitrarily) different relative costs for elective and emergency surgery and for the cost of rescanning. In no case does the choice of strategy change, although, not surprisingly, the magnitude of the cost savings from less frequent recall in strategy C (and hence its net benefit) are somewhat affected by the unit cost of rescans.

Discussion and conclusions

Long-term modelling of the implications of screening programmes is always difficult and typically involves unverifiable assumptions and estimates that have long-term implications. Screening for AAA is an unusual case in that we have been able to compare results from a long-term model initially based on 4-year data with 10 years of observation. The resultant problem was to identify a set of model parameters that would replicate the observed results at 10 years. We were able to do this imperfectly, but nevertheless can have more confidence in the recalibrated model than would otherwise have been appropriate with the original model. Then, with the systematic reviews and the RESCAN analyses of growth and rupture rates, we were able to go further in estimating the cost-effectiveness implications of alternative surveillance strategies.

TABLE 24 Adapted economic model cost-effectiveness results (30-year results) for alternative recall strategies

	Strategy A (1 year, 3 months)	Strategy B (2 years, 6 months)	ICER B compared with A	Strategy C (2 years, 3 months)	ICER C compared with A	Strategy D (1 year, 6 months)
Control arm						
Life-years	12.7157	12.7155		12.7157		12.7156
Cost (£)	271.65	260.23		265.87		265.46
Invited arm						
Life-years	12.7244	12.7240		12.7242		12.7241
Cost (£)	319.89	305.03		311.54		312.20
Difference						
Life-years	0.008659	0.008502	−0.0001578	0.008580	−0.00007983	0.008569
QALYS	0.006754	0.006631	−0.0001231	0.006692	−0.00006226	0.006683
Cost (£)	48.25	44.80	−3.45	45.68	−2.57	46.74
ICER £ per LY	5572	5270	21,853	5324	32,236	5454
ICER £ per QALY	7143	6756	28,016	6825	41,329	6993
Net benefit (λ = £20,000 per QALY)			0.99		1.33	

Costs based on 2010–11 prices – costs and mortality discounted at 3.5%. Survival based on deaths related to AAA deaths, accounting for other causes of death.

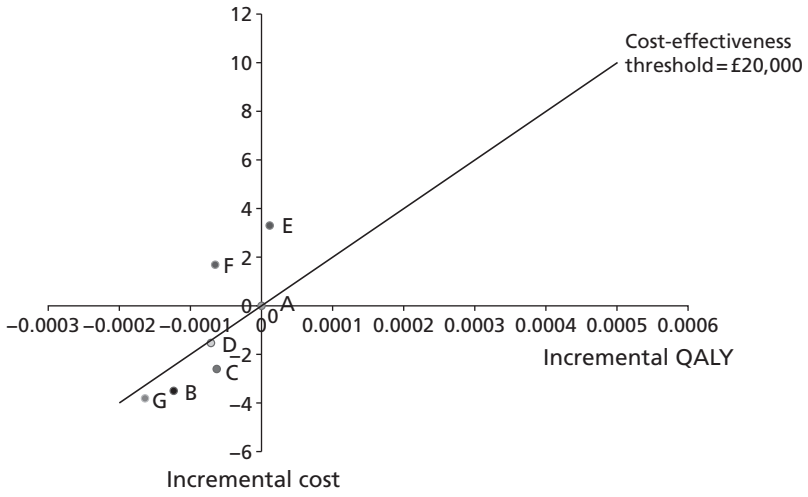


FIGURE 22 Cost-effectiveness plane for new screening interval strategies.

ICER D compared with A	Strategy E (6 months, 3 months)	ICER E compared with A	Strategy F (6 months, 6 months)	ICER F compared with A	Strategy G (3 years, 3 months)	ICER G compared with A
	12.7158		12.7156		12.7155	
	279.92		273.40		262.17	
	12.7244		12.7241		12.7240	
	331.49		323.35		306.64	
–0.00009086	0.008674	0.00001455	0.008576	–0.00008350	0.008449	–0.0002103
–0.00007087	0.006766	0.00001135	0.006689	–0.00006513	0.006590	–0.0001640
–1.51	51.57	3.32	49.95	1.70	44.47	–3.78
16,650	5945	228,111	5824	–20,365	5263	17,972
21,346	7622	292,450	7467	–26,109	6748	23,041
0.10		–3.09		–3.00		0.50

These show that from a cost-effectiveness perspective lengthening the surveillance interval for aneurysms of 4.5–5.4 cm reduces net monetary benefit and we have argued that decreasing that interval would not be practical. However, increasing the interval for recall of men with aneurysms between 3.0 and 4.4 cm from 1 year to 2 years improves cost-effectiveness, but increasing it further to 3 years worsens cost-effectiveness compared with 2 years and (marginally) compared with the current 1-year interval.

It is important to recognise that the absolute differences in outcomes and costs between the surveillance options we have considered are small, as are the absolute values of the net monetary benefit per man invited. They can be put into perspective by multiplying them up to reflect the impact on a large-scale screening programme. With a programme inviting around 260,000 men per year as will broadly be the situation when the NHS programme covers the whole of England,[73] the (present value of the) cost difference for the preferred option would be of the order of £660,000 per year, but the QALY loss would be equivalent to around 16 QALYs. Given the remaining uncertainties there has to be a question of whether or not the differences involved justify a change from the existing surveillance programme particularly as it would have to be explicit that this strategy was expected to be slightly inferior in terms of clinical effectiveness.

TABLE 25 Adapted economic model cost-effectiveness results (30-year results): sensitivity analyses of growth and ruptures rates, operative mortality rates, dropout rates, costs of surgery and cost of rescans. Incremental net benefit (INB) of each strategy is given compared with strategy A (1 year, 3 months)

	Strategy C (2 years, 3 months), INB (£)	Strategy B (2 years, 6 months), INB (£)	Strategy G (3 years, 3 months), INB (£)	Strategy D (1 year, 6 months), INB (£)	Strategy F (6 months, 6 months), INB (£)	Strategy E (6 months, 3 months), INB (£)
Base case	1.33	0.99	0.50	0.10	−3.00	−3.09
Growth and rupture rates from three UK population screening studies[10,37,53] that used internal aortic diameter measurements						
Three UK screening studies	1.09	0.61	−0.11	−0.04	−3.01	−2.96
Growth rates						
Growth rates ↓10%	1.52	1.29	0.94	0.23	−2.95	−3.19
Growth rates ↑10%	1.13	0.68	0.03	−0.04	−3.05	−3.00
Rupture rates						
Rupture rates ↓30%	1.33	0.96	0.47	0.08	−3.02	−3.10
Rupture rates ↑30%	1.33	1.01	0.53	0.11	−2.99	−3.09
Operative mortality rate – base case[a] 0.029 and 0.074						
0.024 (NAAASP)	1.36	1.16	0.58	0.25	−2.85	−3.10
Dropout rate from rescanning – base case 1.4%						
Dropout rate ↓20%	1.47	1.20	0.73	0.20	−3.06	−3.26
Dropout rate ↑20%	1.20	0.80	0.30	0.01	−2.95	−2.94
Costs of elective and emergency surgery – base case £19,985 and £12,806						
Elective ↑10% Emergency ↓10%	1.38	1.07	0.63	0.14	−2.96	−3.10
Elective ↑30% Emergency ↓30%	1.47	1.25	0.88	0.24	−2.87	−3.12
Elective ↓10% Emergency ↑10%	1.28	0.90	0.37	0.05	−3.05	−3.08
Elective ↓30% Emergency ↑30%	1.18	0.72	0.11	−0.05	−3.13	−3.07
Cost of rescanning – base case £68						
Cost of rescan ↓10%	1.10	0.69	0.20	−0.03	−2.81	−2.77
Cost of rescan ↓30%	0.64	0.11	−0.41	−0.28	−2.42	−2.11
Cost of rescan ↑10%	1.56	1.28	0.80	0.22	−3.20	−3.42
Cost of rescan ↑30%	2.02	1.87	1.41	0.47	−3.59	−4.07

↓, decreased by; ↑, increased by.

a In the base case, 0.029 relates to the mortality following surgery for screen-detected individuals, whereas 0.074 relates to those opportunistically detected. The data available from the NAAASP do not make a distinction.

APPENDIX C: Paper 3

Paper 3: Lessons from AAA modelling

Glover MJ, Jones E, Masconi KL, et al. Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening. Med Decis Mak. 2018;38(4):439-451. doi:10.1177/0272989X17753380.
https://doi:10.1177/0272989X17753380

Original Article

Medical Decision Making

Medical Decision Making
2018, Vol. 38(4) 439–451

DOI: 10.1177/0272989X17753380
journals.sagepub.com/home/mdm
SAGE

Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening

Matthew J. Glover*, Edmund Jones*, Katya L. Masconi, Michael J. Sweeting, Simon G. Thompson, and SWAN Collaborators

Abstract

Markov models are often used to evaluate the cost-effectiveness of new healthcare interventions but they are sometimes not flexible enough to allow accurate modeling or investigation of alternative scenarios and policies. A Markov model previously demonstrated that a one-off invitation to screening for abdominal aortic aneurysm (AAA) for men aged 65 y in the UK and subsequent follow-up of identified AAAs was likely to be highly cost-effective at thresholds commonly adopted in the UK (£20,000 to £30,000 per quality adjusted life-year). However, new evidence has emerged and the decision problem has evolved to include exploration of the circumstances under which AAA screening may be cost-effective, which the Markov model is not easily able to address. A new model to handle this more complex decision problem was needed, and the case of AAA screening thus provides an illustration of the relative merits of Markov models and discrete event simulation (DES) models. An individual-level DES model was built using the R programming language to reflect possible events and pathways of individuals invited to screening v. those not invited. The model was validated against key events and cost-effectiveness, as observed in a large, randomized trial. Different screening protocol scenarios were investigated to demonstrate the flexibility of the DES. The case of AAA screening highlights the benefits of DES, particularly in the context of screening studies.

Keywords

abdominal aortic aneurysm, decision analytic model, discrete event simulation, Markov model, screening

Received February 3, 2017; accepted December 15, 2017.

Modeling will almost always constitute an essential component of an economic evaluation to inform decision making, to overcome the limitations of available randomized trial data.[1] In screening studies, where much of the cost is upfront and benefits accrue over a long period of time, there is a need for modeling approaches that can contribute to long-term economic evaluations. The choice of modeling technique is at the discretion of the analyst and often reflects an implied trade-off between simplicity and realism in reflecting a disease's natural history, treatment, and patient outcomes.[2,3] Markov models have been widely used, as they provide a simple mechanism to estimate the long-term costs and effects of healthcare interventions.[4] Discrete event simulation

Health Economics Research Group, Brunel University London, Uxbridge, Middlesex, UK (MJG); Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, Cambridgeshire, UK (EJ, KLM, MJS, SGT); SWAN collaborators (see Acknowledgements). Financial support for this study was provided by the UK National Institute for Health Research (NIHR) Health Technology Appraisal (HTA) programme (project number 14/179/01). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, National Health Service or the Department of Health. Work done at the University of Cambridge was additionally funded by the UK Medical Research Council (MR/L003120/1), the British Heart Foundation (RG/13/13/30194), and the UK National Institute for Health Research (Cambridge Biomedical Research Centre)

*These authors contributed equally to this work.

Corresponding Author

Matthew J. Glover, Health Economics Research Group, Brunel University London, Uxbridge, Middlesex, UB8 3PH, UK; +44 (189) 526-7835. (matthew.glover@brunel.ac.uk)

(DES) is a less common alternative that avoids the use of states and fixed cycle lengths and instead models events at the individual level. Some have suggested that DES should always be preferred, whereas others have highlighted particular circumstances where DES should be favoured.[5,6] The case of abdominal aortic aneurysm (AAA) screening offers an illustration of the relative merits of these techniques.

An AAA is commonly defined as an aortic diameter $\geq$3.0 cm. A long-term Markov model demonstrated that offering population screening for AAA to men aged 65 y in the UK was likely to be highly cost-effective.[7] This model was largely populated using data from the 4-y follow-up of the randomized Multicentre Aneurysm Screening Study (MASS),[8] and adopted the same screening methods, surveillance intervals (for 3.0 to 5.4 cm AAAs) and AAA diameter threshold (5.5 cm) for referral for elective surgical intervention as in MASS. The MASS trial of 67,800 men aged 65 to 74 showed that an invitation to a one-off ultrasound scan and surveillance or referral for elective surgical intervention of identified AAAs was effective in reducing AAA-related mortality, initially at the 4-y follow-up,[8] and subsequently at longer-term follow-up.[9] The results from the modeling and the MASS trial formed a large part of the evidence base supporting the phased implementation from 2009 of the NHS AAA Screening Programme (NAAASP) in England, with full coverage across the UK by the end of 2013.

Research into the clinical and cost-effectiveness of AAA screening has evolved since the first Markov modeling was performed, with the emergence of new data and evidence. Initial observational data from NAAASP suggested that the current prevalence of AAAs is substantially lower than that observed in MASS (1.6% v. 4.9%). The Markov model (MM) was updated to reflect this lower prevalence as well as changes in costs, the increased use of endovascular surgical techniques, meta-analyzed data on growth and rupture rates,[10,11,11] and longer-term MASS follow-up.[12] The results suggested screening is still likely to be highly cost-effective, with a long-term incremental cost-effectiveness ratio of £7,370 (95%CI, £5,467 to £9,443) per quality-adjusted life year (QALY). Other studies based on populations in Denmark[13] and Sweden,[14] with similar AAA prevalence, support this conclusion.

However, different programs and randomized trials have adopted diverse surveillance intervals, with little consensus on optimal intervals.[15] More substantial surveillance data from the RESCAN project was incorporated into an adapted MM to investigate different surveillance intervals, and the results suggested that lengthening the time between rescans for men with the smallest aneurysms could be done at acceptable clinical risk[10] and would be a cost-effective strategy.[11] Some of the protocols around screening have also come under scrutiny. For example, the definition of an AAA as an aortic diameter $\geq$3.0 cm is somewhat arbitrary: there is evidence that many individuals with screen-detectable sub-aneurysmal aortic dilation (2.5 to 2.9 cm) will progress to AAA within 10 y.[16] The implications for screening remain unclear. There have also been some suggestions that the surgical threshold itself should be altered.[17]

In light of these findings, the decision problem no longer relates only to "screening" v. "no screening" for older men but has evolved to include the circumstances under which screening may be cost-effective.[18-21] Modeling allows these questions to be addressed without conducting costly primary research and could extend to varying a number of fixed parameters (e.g., surveillance intervals, the AAA diameter threshold for referral for elective surgery, screening of women, targeted screening based on patient characteristics). However, MMs can be inflexible. This inflexibility was demonstrated by the extensive reprogramming needed to build tunnel states in the MM when the model was adapted to assess different surveillance intervals.[11] Such analyses are important for existing programs aiming to improve their performance or extend population coverage, as well as for other countries considering implementation. Therefore, a model better able to handle this decision problem is required. This paper describes: 1) the development and validation of a DES model to estimate the clinical and cost-effectiveness of AAA screening; and 2) the use of the DES to explore the cost-effectiveness of screening under various scenarios, which was not possible with the original MM.

Methods

Development of a Simulation Model

A DES was implemented using the freely available statistical programming language R and based on the original MM.[7] The original MM defined several health states that related to AAA identification, aortic diameter ($<$3.0 cm, 3.0 to 4.4 cm, 4.5 to 5.4 cm, $\geq$5.5 cm) and associated events (rupture, surgical consultations, elective and emergency AAA repair, death). A set of transition probabilities determined movements between health states for the two populations (invited to screening and not invited). The MM operates at the cohort level: events, mean costs, and QALYs are calculated from the proportions of the cohort that inhabit the different health states in each 3-

mo cycle. In contrast, the DES functions at an individual level, simulating sequences of events that occur as a continuous process over time and calculating the associated mean costs and QALYs. It allows individual patient heterogeneity to be characterized and accounts for events as they occur, removing the need for any assumptions relating to averaging costs or outcomes across cycles.

An Event Scheduling Approach

Full details of the DES are available in the SWAN project National Institute for Health Research Health Technology Assessment monograph. The DES adopts an event-scheduling approach by generating a sequence of events for each individual, using a list of events that are "scheduled" for the future (future events list; FEL). The DES has an explicit simulation clock, chooses the event that has the earliest sampled time, and records it in the individual's sequence of events. It then schedules, reschedules, or cancels other scheduled events as necessary, updating the FEL (for example, if a surveillance rescan finds that an individual's aortic diameter is above the threshold for elective surgery, then a consultation is scheduled). This process is repeated until death or censoring (dependent on model time horizon). The possible sequences of events are shown in Figure 1.

Individuals are assigned an aortic diameter, drawn from a population distribution, and a latent parameter describing the growth rate of their aorta over time. Details of the aortic growth model are given in the Supplementary Material. Non-AAA death and AAA rupture events are scheduled in the future, and if the individual is in the "invited" group, then an invitation is also scheduled. If the individual is in the "non-invited" group, then an "incidental detection" event is scheduled. The time to AAA rupture is dependent on the individual's initial aortic diameter and their latent growth rate. In most instances, the scheduled AAA rupture time will be so far in the future that there is no chance of the event occurring. The "incidental detection" event is scheduled to occur only after the time at which an individual's aortic diameter reaches the diagnosis threshold (e.g., 3.0 cm).

The DES simulates people in pairs, like identical twins, one of whom is in the invited group and one in the non-invited group. The twins have certain characteristics in common: they have the same times of non-AAA death and AAA rupture in their FELs, the same initial aortic diameter and growth rate, and the same values of certain parameters, such as indicators (binary variables) for whether they would be contraindicated for surgery and whether they would survive emergency surgery. Differences in costs and outcomes result from the different events that are scheduled due to involvement or otherwise in the screening programme.

Joint Continuous AAA Growth and Rupture Model

A major difference between the DES and the MM is that the DES uses a joint continuous-time model for aortic growth and rupture[11] rather than defining 4 AAA size states. Additionally, when an individual's aorta is scanned, the measurement is generated by calculating the diameter according to this model and adding measurement error, which is specific to the type of scan used (i.e., ultrasound or computed tomography [CT]). Further alterations related to the move from fixed cycles to a continuous process were made. For example, surgical waiting time was previously separated into two periods: the time from discovery to consultation (71 d) and from consultation to surgery (59 d).[22] In the MM, it was assumed that this total waiting time could be considered as a 3-mo cycle. The DES would enable these periods to be easily changed if appropriate.

A Hierarchy of Functions

The R program for the DES is made up of a hierarchy of functions or routines: 1) a probabilistic sensitivity analysis (PSA), which consists of running the main analysis multiple times; 2) the main individual patient simulation analysis, which consists of simulating and analyzing multiple pairs of individuals; 3) the function to process one pair; and 4) the function to generate a sequence of events for an individual. These functions are shown in Figure 2.

The R code developed for this project is available on request from the authors.

The DES involves a large number of parameters. These can be classified into several sets: global fixed parameters, global uncertain parameters, and parameters that are specific to an individual or a pair of twins ("global" refers to population parameters and "uncertain" means that a parameter follows a random distribution). Like the functions, these sets form a hierarchy. For example, in a PSA, a beta distribution is used to generate the probability that an individual will die following emergency surgery, if they have emergency surgery. The parameters of the beta distribution are global fixed parameters, and the probability is a global uncertain parameter. In the main analysis, when a pair of twins is created, the probability is used as the parameter in a Bernoulli distribution to generate the indicator for the

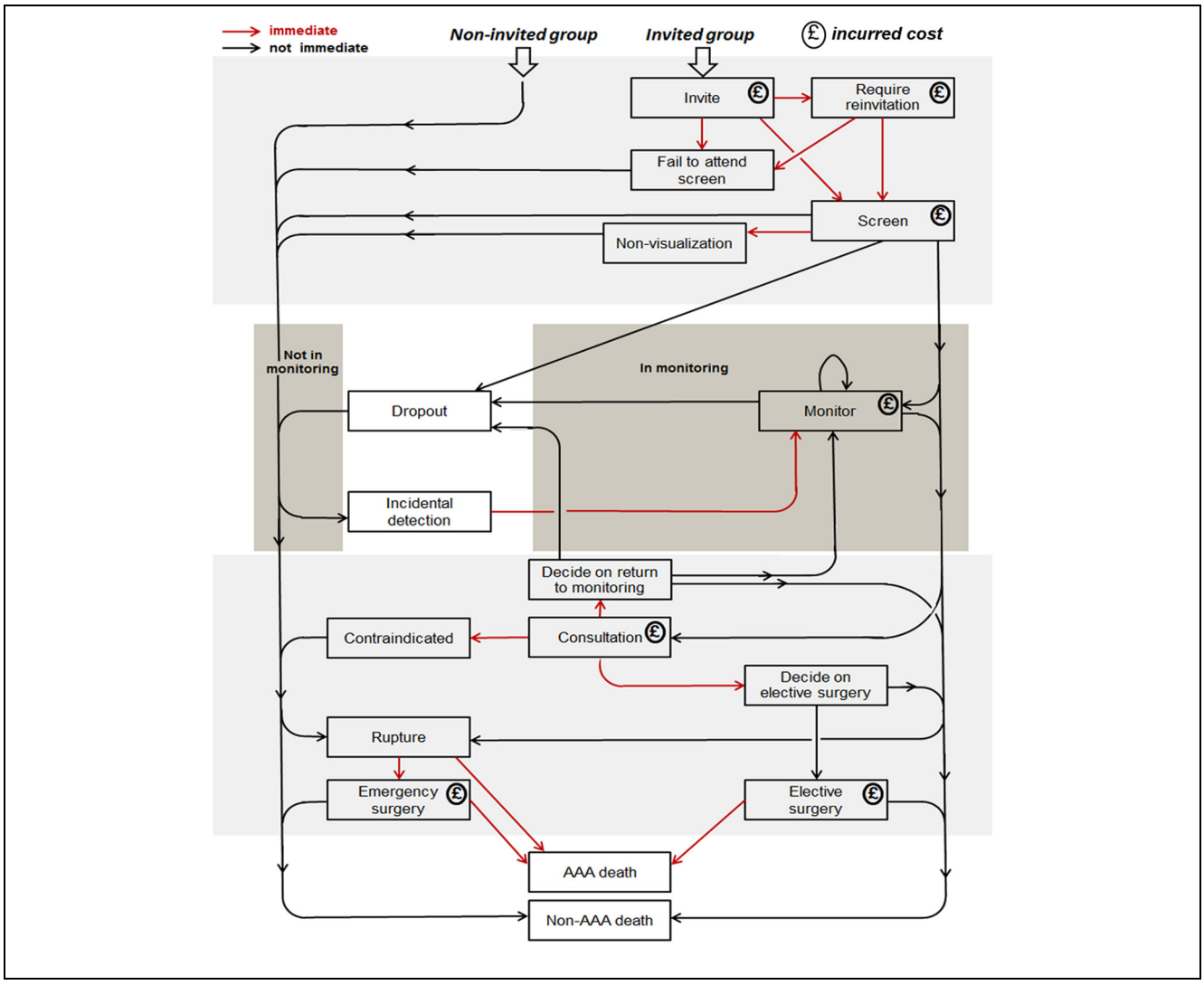


Figure 1 Possible sequences of events in the abdominal aortic aneurysm (AAA) screening discrete event simulation model.

twins' emergency surgery outcomes. The indicator is a variable specific to the pair of twins.

The hierarchy extends downward by two more levels. When the DES generates an individual's sequence of events, it needs to record the intervention group (i.e., invited or not), which can be regarded as an individual-specific parameter, and when an aorta measurement is generated, a new and unique value of the measurement error is created, which can be regarded as an event-specific parameter. Figure 2 shows how the sets of parameters are passed from one function to another. The definitions of the parameters require judgement and depend on the nature of the input data. For instance, the costs of the scans and other events could be defined as global fixed parameters, if their values are known with great certainty, or global uncertain parameters if they are not.

Cost-effectiveness Analysis

The cost-effectiveness analysis consists of simulating a large number of individuals, calculating their life-years and costs, and calculating the mean life-years and costs over all the patients in the groups invited to screening and not invited. Given that the model outputs are driven by those individuals who have an AAA, these individuals were oversampled (and later calculations were adjusted to account for this), which reduces considerably the Monte Carlo error when estimating incremental effects and costs. PSA is conducted to account for uncertainty in

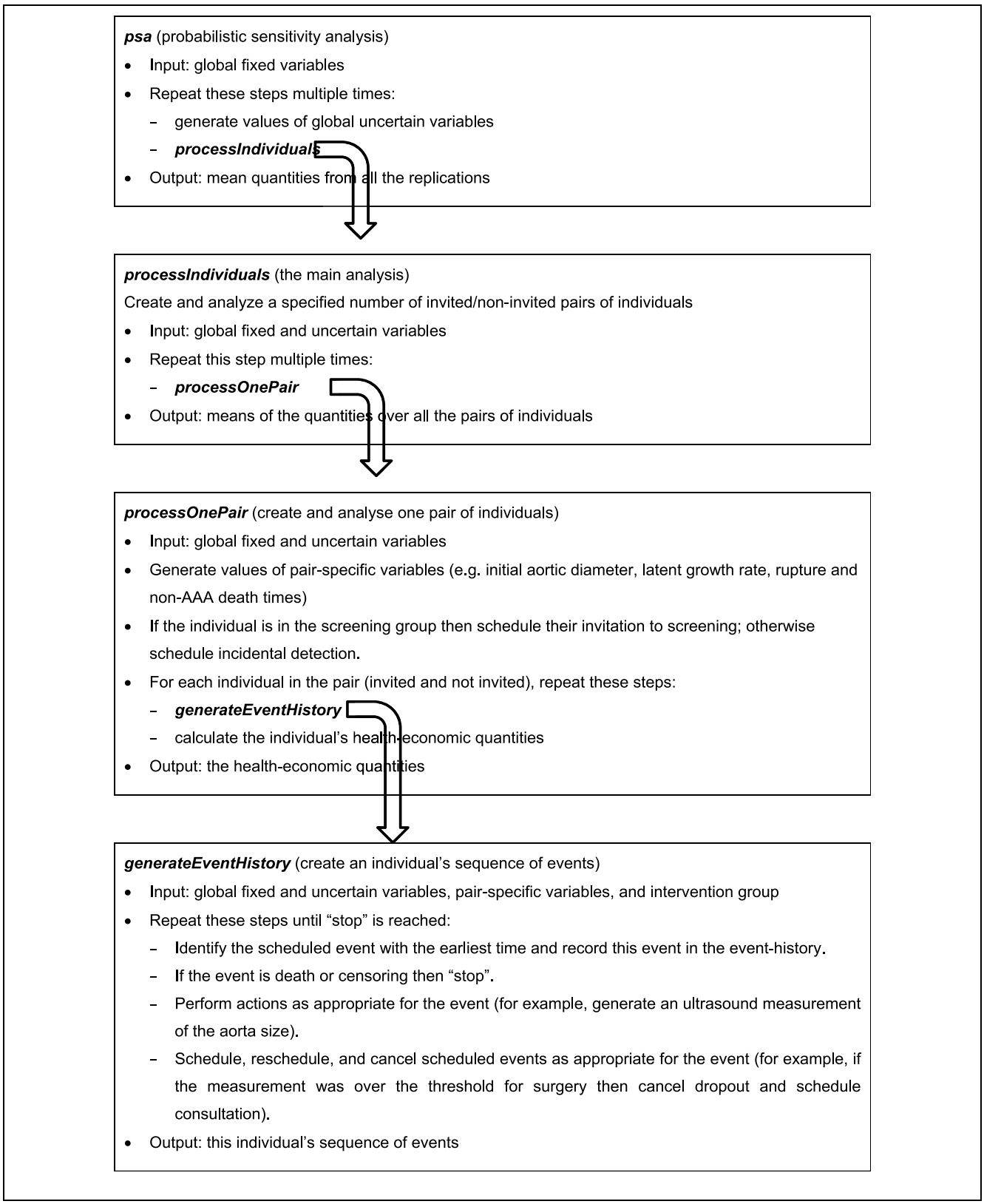


Figure 2 Abdominal aortic aneurysm (AAA) screening discrete event simulation model: Hierarchy of functions.

the model parameters; repeated sets of values for the global uncertain parameters are generated, and the main analysis is run for each set of values. The mean incremental cost and effectiveness (i.e., QALYs), together with the incremental cost-effectiveness ratio (ICER) and incremental net monetary benefit (INMB), are calculated for each set, and the distribution of values is used to estimate the probability that the screening program is cost-effective.

Programming Practice

The DES is a moderately complicated computer program, and it was therefore necessary to follow basic principles of good programming to ensure that it would run correctly and be maintainable and usable in the future. For example, each function has a clearly defined single purpose that can easily be understood from its name, and priority was given to making the source-code simple, clear, and readable (by other people) rather than computationally fast. The DES was written using a mostly "functional" programming style: the basic building-blocks are functions, and functions do not modify things outside themselves but simply perform actions and then either display output or return relevant quantities (e.g., parameters, event-times). R is convenient for statistical and scientific programming, and it allows loops (or iterative processes) to be written to run in parallel (which is not the case with all programming languages). The DES is eminently suited to parallelization, which speeds it up considerably.

Model Validation

The original MM was validated against the MASS trial 4-y follow-up to check the appropriateness of model outputs. The validation involved comparing the numbers of key events (e.g., AAA ruptures, number of elective operations) and mean costs and life-years, as observed in the trial, with model outputs based on a simulated population of the same size. The MM was able to replicate the observed data reasonably.[7,23] For the DES, a similar process was carried out, again using the 4-y MASS follow-up data. Costs and life-years were discounted at 6% and 1.5% per y, respectively, to be consistent with the original rates used in the 4-y follow-up analysis.

Input parameters for the DES were derived from the MASS 4-y follow-up, where possible, including non-AAA death rates, to enable validation. Other adaptations were made to improve the DES model fit to the observed MASS outputs. The parameters for the aorta growth model were chosen such that, at baseline, there were the same proportions of individuals with aortic size <3.0 cm, 3.0 to 4.4 cm, 4.5 to 5.4 cm and ≥ 5.5 cm as in MASS. Growth rates were based directly on those observed in the screen-detected MASS population, with growth rates for those 2.0 to 2.9 cm extrapolated from a fitted mixed model, and growth rates set to zero for those <2.0 cm at baseline (see Supplementary Material for more details). All aorta measurements that were performed by CT scan (at consultation only) were, on average, 0.24 cm larger than an ultrasound scan, to account for CT scanning measuring outer-to-outer rather than inner-to-inner diameters.[11] Individuals were censored at uniformly random times between 3 and 5.25 y, because the "4-y" follow-up of the MASS data had censoring times similar to this uniform distribution. Full details of the input parameters and characterization of uncertainty is detailed in Supplementary Table 1.

New Model Scenarios

After validation against the MASS trial 4-y follow-up, parameter values in the DES were updated to reflect more contemporaneous estimates, the full details of which are provided in Supplementary Table 2. National mortality statistics[24] were used for non-AAA death rates. The NAAASP baseline aortic diameter distribution was used in the aorta growth and rupture model, with growth and rupture rates based on RESCAN data.[25] As before, growth rates <3.0 cm were extrapolated from a model or were set to zero. Costs, attendance rate, and other parameters were updated as described by Glover and others,[12] with QALYs estimated by applying population norm utility weights to life-years accrued. The model structure was further altered to allow for endovascular aneurysm repair (EVAR) as well as open repair, with a proportion of surgery by EVAR, which incurred a different cost and post-operative mortality rate compared to open surgery.[11] The base case was run over a 30-y time horizon for 65-y-old men invited or not invited to screening, for 10 million pairs of individuals. PSA, based on 1,000 runs with 500,000 pairs of individuals, was used to characterize uncertainty in input parameters. Costs and QALYs were discounted at 3.5% per y.

Two modeling scenarios were explored, showing the flexibility of the DES to estimate the cost-effectiveness of screening under various protocols. The first built on an analysis previously performed using the MM, to identify more cost-effective surveillance intervals for men in the screening program.[11] The second allowed the inclusion of surveillance for men with sub-aneurysmal aortic diameters (2.5 to 2.9 cm at first screen). Each of these different scenarios was compared to the existing program in terms of costs and QALYs.

Table 1 Life-years and Costs According to the 4-y MASS Follow-up, Markov Model (Kim and others[7]) and the DES[a]

	MASS Observed	Markov Model	DES Model
Non-invited group			
Life-years	3.816	3.905	3.753
Cost	£35.03	£32.74	£39.11
Invited group			
Life-years	3.819	3.907	3.754
Cost	£98.42	£98.32	£101.97
Difference			
Life-years	0.0022	0.0017	0.0015
Cost	£63.39	£65.58	£62.86
ICER	£28,400	£37,700	£42,137
(95% CI)	(£15,000, £146,000)	(£19,700, £147,000)	(£19,935, £3,277,596)[b]

DES, discrete event simulation; ICER, incremental cost-effectiveness ratio; MASS, Multicentre Aneurysm Screening Study.
[a]Life-years discounted at 1.5% per y and costs at 6% per y.
[b]Reported as uncertainty interval produced by 1,000 probabilistic sensitivity analysis (PSA) iterations (after assigning ICERs with negative incremental effects and positive costs to be infinite). Mean estimates from 1,000 PSA iterations for the difference in life-years, costs, and the ICER were 0.0015, £62.91 and £46,032, respectively.

Scenario 1: Different Surveillance Intervals. Analysis performed using the MM suggested that lengthening surveillance intervals for the smallest identified AAAs would be cost-effective according to thresholds commonly adopted in the UK (£20,000 to £30,000 per QALY).[11] However, the different surveillance strategies considered were limited to varying the time between monitoring for two AAA size groups (3.0 to 4.4 cm, and 4.5 to 5.4 cm). Unlike the MM, the DES can be easily adapted to use any number of plausible AAA size cut-offs, or to consider differing surveillance intervals. Here, the current NAAASP surveillance strategy of 1-year (3.0 to 4.4 cm AAAs) and 3-month (4.5 to 5.4 cm AAAs) intervals is compared to a strategy of 2-year (3.0 to 3.9 cm AAAs), 1-year (4.0 to 4.4 cm AAAs), and 3-month (4.5 to 5.4 cm AAAs) intervals.

Scenario 2: Inclusion of Sub-aneurysmal (2.5 to 2.9 cm) Aortas. The threshold definition of an AAA used in the DES was lowered from 3.0 cm to 2.5 cm. Individuals identified with an aortic diameter of 2.5 to 2.9 cm had surveillance scans scheduled at intervals of 5 years, with the intervals currently adopted by NAAASP maintained for AAAs between 3.0 cm and 5.4 cm.

Results

DES Model Validation

The DES validated reasonably against the MASS 4-y data. The DES broadly agreed with both the observed MASS 4-y follow-up and the original MM in terms of differences in life-years and costs (Table 1). However, like the MM, the 4-y ICER for the DES was higher than in the MASS data. Table 2 shows the total numbers of events as observed in MASS and estimated by the original MM and the DES. For most events, the numbers of events were similar, as were the ratios of events in the DES to events in MASS show. The numbers of non-AAA deaths matched very closely, primarily because the MM and DES used non-AAA death rates from MASS and most individuals do not experience AAA rupture. As examples of cumulative events over time, Figure 3 shows the numbers of emergency operations in the non-invited group and AAA deaths in the invited group estimated by the DES, compared to the observed numbers in the MASS 4-y follow-up.

New Model Scenarios

The updated model after validation, using contemporaneous data sources, estimated a 30-y ICER of £6,352 (95%CI, £5,059 to £8,808) per QALY (Table 3). This compares to a 30-y ICER of £7,370 produced by the MM.[12] The 1,000 iterations on the cost-effectiveness plane and cost-effectiveness acceptability curve shown in Figure 4 demonstrate that a one-off invitation to AAA screening and subsequent follow-up of identified AAAs is highly likely to be cost-effective, with no iterations outside the cost-effective region.

The estimated INMBs for both new scenarios were positive when compared with the existing screening program at a willingness-to-pay threshold of £20,000 per QALY (Table 4). The longer surveillance interval for the

Table 2 Key Events Observed in the MASS 4-y Follow-up, and as Estimated by the Markov Model (Kim and others[7]) and the DES

	MASS Observed	Markov Model[a]	DES Model[a]	DES Model (% of MASS)
No invitation group				
Elective operation	100	83	98	98
Emergency operation	62	62	68	110
Rupture	138	141	154	112
Contraindicated for elective surgery	NA	14	16	NA
AAA death	113	109	120	106
Non-AAA death	3,750	3,724	3,696	99
Invited group				
Elective operation				
Resulting from screen detection	295	282	330	112
Resulting from incidental detection	31	25	27	86
Emergency operation	28	34	30	106
Rupture	66	78	67	102
Contraindicated for elective surgery				
Resulting from screen detection	41	46	54	131
Resulting for incidental detection	NA	5	5	NA
AAA death	65	69	63	98
Non-AAA death	3,694	3,724	3,700	100
Loss to recall follow-up	290	289	278	96

AAA, abdominal aortic aneurysm; DES, discrete event simulation; MASS, Multicentre Aneurysm Screening Study; NA, not available.
[a]Estimated for a sample size of 33,961 participants in the control group and 33,839 in the invited group, as in MASS.

smallest AAAs (scenario 1) resulted in a small cost saving, as those with 3.0 to 3.9cm AAAs are screened less often than in the existing program. The longer surveillance interval was also associated with almost no change in QALYs. Extending the surveillance program to those with 2.5 to 2.9 cm aortic diameters (scenario 2) was associated with additional benefits but greater costs. However, the model suggests that rescanning these individuals at 5-y intervals could be cost-effective, with an INMB of £10 per individual invited to screening. The estimate of the INMB for scenario 1 was positive; however, the CI from the PSA included zero; the probability of it being cost-effective compared with the current strategy was 0.68. The INMB for scenario 2, with a CI that excludes zero, suggests that a surveillance for those with an aortic diameter between 2.5 and 2.9 cm is a cost-effective strategy as compared with the current strategy.

Discussion

In assessing screening programs, modeling techniques are particularly relevant given that most of the costs are upfront, but benefits continue to accrue over time. The validated MM built by Kim and others[7] demonstrated that AAA screening for men aged 65 y in the UK was likely to be cost-effective, but the model was inflexible when trying to address questions around configuration and optimization of screening. The creation of the DES has overcome these problems and the case of AAA screening highlights situations where DES may provide the most appropriate method to perform an economic evaluation, particularly when surveillance or rescreening is based on patient characteristics or risk markers (in this instance, AAA size). However, it cannot be asserted that the decision around the conceptual model should have been different at the onset of the research. The evolution of the decision problem has necessitated the re-conceptualization. Indeed, the MM served a valuable and timely purpose in showing the long-term cost-effectiveness of a one-off invitation to screening in the UK for men aged 65 y.

The DES built was based on the original MM, which itself was largely based on the MASS trial. This allowed the validation of the DES against observed MASS follow-up events and cost-effectiveness results. Overall, the DES validation process was similar to that for the MM, and no major changes were necessary to produce a comparison of key outputs. Given that some of the parameters in the model were not based on data directly observed in MASS (e.g., incidental detection rate), further calibration to fit MASS-observed data could have been undertaken to better replicate the number of events or the cost-effectiveness results based on trial data. This type of calibration was performed to produce a better fit using the MM.[24] However, when trying to validate a model against data from a study

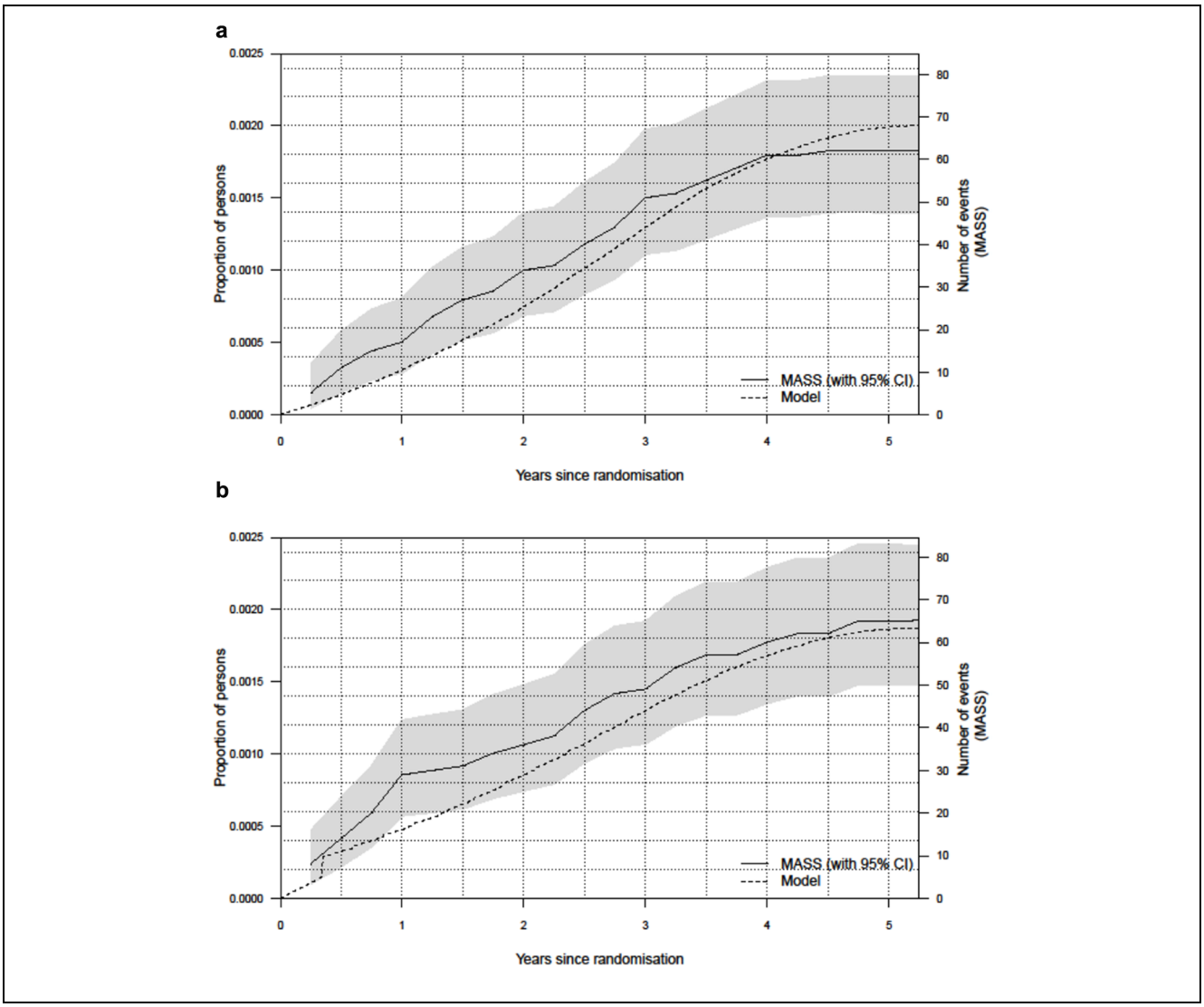


Figure 3 Cumulative numbers of events in the 4-y MASS data and the DES for: (a) emergency operations in the non-invited group and (b) AAA deaths in the invited group. AAA, abdominal aortic aneurysm; DES, discrete event simulation; MASS, Multicentre Aneurysm Screening Study.

in this way, there is a risk of creating a cyclical process. If too much information from the study is used, then the model output might match the data very well but the model may not predict well over a longer time horizon, rather like the issue of overfitting in statistical modeling. The differences between the DES and MM may be partly explained by the approach developed to handle aortic growth of those AAA <3.0 cm at first screen. The performance of the DES in the validation gives some confidence in using the model to extrapolate over a longer term.[12,23]

The general advantages of DES in health economic modeling have been extolled previously.[2,6,26–28] They offer a decision modeler greater flexibility to adequately reflect clinical pathways, characterize baseline patient heterogeneity, allow event rates that change over time or depend on patient characteristics, and avoid the constraints of state transitions and fixed cycles imposed by the MM. In the case of AAA screening, there are 3 particular characteristics that mean that a DES is superior: firstly being able to define the size of a AAA as a continuous variable, which also allows measurement error in the ultrasound observations; secondly, allowing heterogeneity in the AAA growth rates between different individuals, with uncertainty easily characterised, something

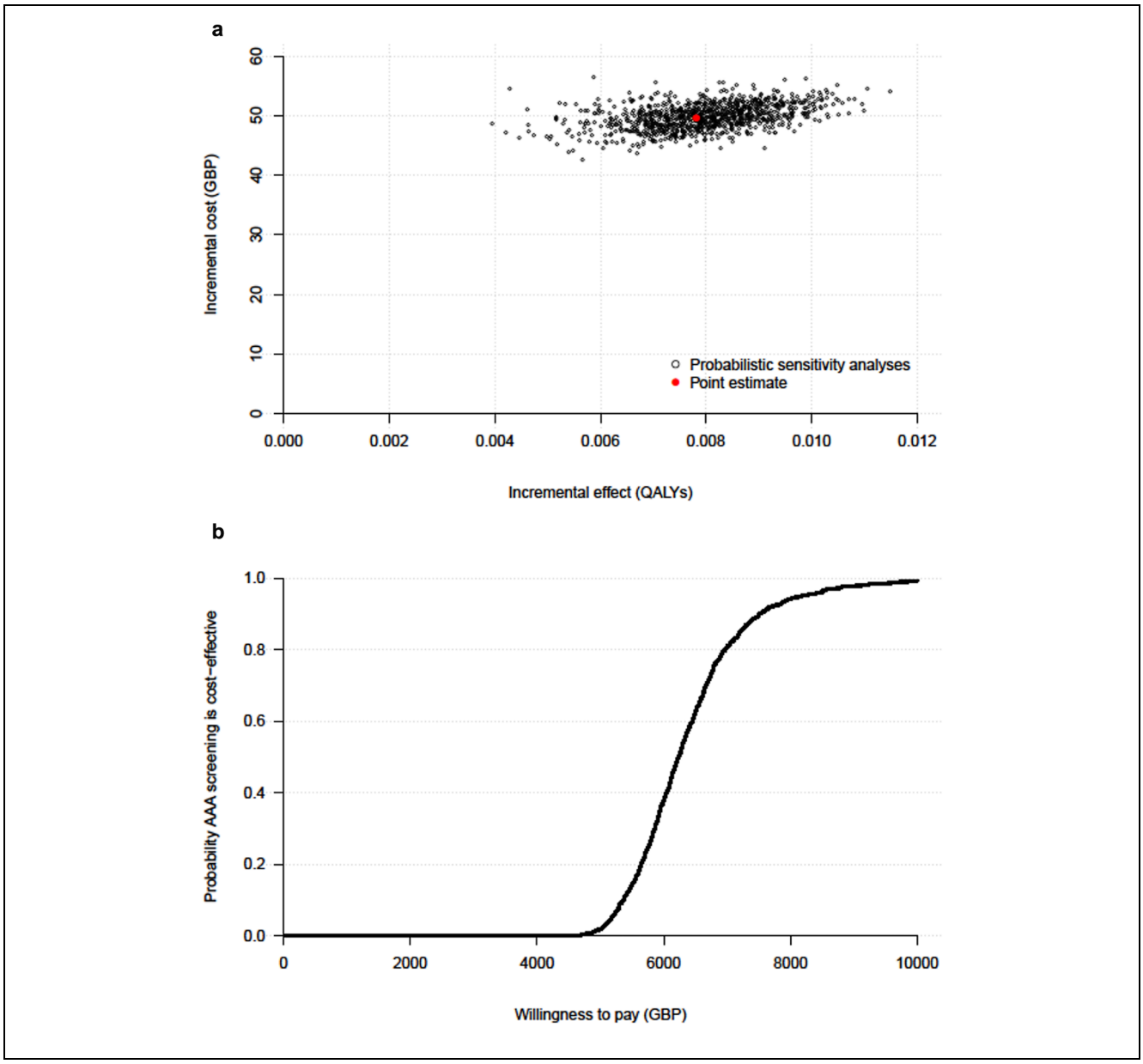


Figure 4 Long-term (30-y) cost-effectiveness of one-off invitation to AAA screening: (a) 1,000 probabilistic sensitivity analysis iterations (current NAAASP program), (b) cost-effectiveness acceptability curve. AAA, abdominal aortic aneurysm; NAAASP, National Health Service AAA screening programme .

that is difficult to recreate in an aggregate discrete-state MM, even by varying the transition probabilities over time; and thirdly the ease with which time-varying surveillance intervals and other changes to the screening programme can be defined and evaluated. There are generally perceived trade-offs between the simplicity of an MM and the complexity of a DES, particularly related to model build time, potential data requirements, and model run times; the latter will always be a consideration. However, the advantages of DES models start to outweigh other factors as the complexity of the decision problem and modeled pathway increases, especially if structural modifications and further data analysis are necessary to deal with different scenarios.

The results of the scenarios presented here would be difficult to replicate in an MM, constrained by the state

Table 3 Discrete Event Simulation: Long Term (30-y) Cost-effectiveness of One-off Invitation to AAA Screening for 65-y-old Men[a]

	DES Model
No invitation group	
Life-years	12.601
QALYs	9.681
Cost	£164
Invited group	
Life-years	12.611
QALYs	9.689
Cost	£213
Difference	
Life-years	0.01031
QALYs	0.00781
Cost	£50
ICER (QALYs)	£6,352
(95%CI)[b]	(£5,059 to £8,808)

DES, discrete event simulation; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.
[a]Life-years, QALYs, and costs discounted at 3.5% per y.
[b]Reported as uncertainty interval produced by 1,000 probabilistic sensitivity analysis iterations. Mean estimates from 1,000 PSA iterations for the difference in life-years, QALYs, costs, and the ICER were 0.01050, 0.00796, £50 and £6,388, respectively.

transition approach. The problem of modeling different surveillance intervals was encountered in previous work on AAA screening and required the re-programming of 6 different incarnations of the MM,[11] each containing a different number of 3-mo tunnel states. For each surveillance interval that was different from the original screening strategy, tunnel states that allowed unobserved aortic growth and a related rupture rate in each cycle were necessary. Members of the cohort in these unobserved tunnel states were then able to move back into observed states at each rescan. It would have been desirable to explore more combinations of screening intervals and associated AAA size cut-offs. However, the structural changes that are a necessity in an MM would have made this a time-consuming process, thus limiting the number of analyses that could be considered. Conversely, the DES can easily assess any combination of surveillance intervals. To change the interval for patients with a 3.0 to 3.9 cm AAA from one to two years is trivial, because the DES is programmed to allow the input of any chosen partition of the aortic size range with an associated screening interval for each part. This is only possible because an individual's AAA size is measured on a continuous scale. The problems of state transition are further demonstrated when trying to assess the cost-effectiveness of including sub-aneurysmal AAAs in a screening program. In the MM, a new AAA state (2.5 to 2.9 cm with 5-y surveillance intervals) would need to be incorporated, with extensive reprogramming. In the DES, all that is necessary is to insert "2.5" in the list of surveillance thresholds and "5" in the list of intervals. In addition, the DES parameters can also be easily made to depend on individual-level covariates (e.g., age-dependent mortality rates after surgery).

The DES can be used to assess the cost-effectiveness of other policy-relevant protocol changes, including the surgical threshold, the age at first screen, and recalling all those screened normal at first screen after a period of time. The model has been parametrized as part of the SWAN[29] study

Table 4 Long-term (30-y) Cost-effectiveness. Scenario 1: Surveillance Intervals of 2 Y (3.0–3.9 cm AAAs), 1 Y (4.0–4.4 cm AAAs) and 3 Mo (4.5–5.4 cm AAAs). Scenario 2: Inclusion of Sub-aneurysmal (2.5–2.9 cm) AAAs in Screening Programme[a]

	Current Strategy	Scenario 1	Scenario 2
Mean incremental QALYs	0.00781	0.00781	0.00860
Mean incremental cost	£49.61	£48.54	£55.17
Compared to current strategy:			
Mean incremental QALYs	NA	0.00000	0.00080
Mean incremental cost	NA	-£1.07	£5.56
ICER (QALYs)	NA	Dominant	£7,002
(95%CI)[b]			(4,615 to 12,233)
INMB[c]	NA	£0.99	£10.33
(95%CI)[b]		(-2.03 to 3.35)	(2.99 to 21.52)

ICER, incremental cost-effectiveness ratio; INMB, incremental net monetary benefit; QALY, quality-adjusted life year.
[a]Life-years, QALYs and costs discounted at 3.5% per y.
[b]At a willingness-to-pay of £20,000 per QALY.
[c]Reported as uncertainty interval produced by 1000 probabilistic sensitivity analysis iterations Mean estimates from 1000 PSA iterations for the difference in QALYs, costs, and the ICER for scenario 1 were 0.00000, £−1.08, and NA; and for scenario 2 were 0.00080, £5.58 and £7,233, respectively.

and used to assess the likelihood of screening (with various protocols) women being cost-effective.

The restructuring of the model as a DES was, as might be expected, a relatively complex undertaking. Nevertheless, coding in a language such as R enables greater clarity and transparency compared to software designed for simulation modeling. However, the computational requirements of the DES were extensive, given the number of individuals needed to reduce sampling variation to an acceptable level and characterizing uncertainty through PSA. Run time was in the region of 24 h to run the model with 500,000 patients and 1,000 PSA iterations, even with parallelization and the use of a high-powered computer.

Acknowledgments

SWAN collaborative group: Simon G. Thompson (Chief Investigator), Michael J. Sweeting, Edmund Jones, Katya L. Masconi (University of Cambridge); Janet T. Powell, Pinar Ulug (Imperial College London); Matthew J. Bown (University of Leicester); Matthew J. Glover (Brunel University London). The authors thank colleagues at Brunel University London and in particular Professor Doug Coyle and Professor Martin Buxton for comments on versions of this manuscript. Any remaining deficiencies or inaccuracies are solely the responsibility of the authors.

Supplementary Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

ORCID iD

Matthew J. Glover http://orcid.org/0000-0001-9454-2668

References

1. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in economic evaluation: An unavoidable fact of life. *Health Econ*. 1997;6:217–27.
2. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15:1295–310.
3. Halpern MT, Luce BR, Brown RE, Geneste B. Health and economic outcomes modeling practices: A suggested framework. *Value Health*. 1998;1:131–47.
4. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*. 1998;13:397–409.
5. Karnon J, Haji Ali, Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics*. 2014;32:547–58.
6. Caro JJ, Moller J, Getsios D. Discrete event simulation: The preferred technique for health economic evaluations? *Value Health*. 2010;13:1056–60.
7. Kim LG, Thompson SG, Briggs AH, Buxton MJ, Campbell HE. How cost-effective is screening for abdominal aortic aneurysms? *J Med Screen*. 2007;14:46–52.
8. Ashton HA, Buxton MJ, Day NE, et al. The multicentre aneurysm screening study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: A randomised controlled trial. *Lancet*. 2002;360:1531–9.
9. Thompson SG, Ashton HA, Gao L, Buxton MJ, Scott RA, Multicentre Aneurysm Screening Study (MASS) Group. Final follow-up of the multicentre aneurysm screening study (MASS) randomized trial of abdominal aortic aneurysm screening. *Br J Surg*. 2012;99:1649–56.
10. RESCAN Collaborators, Bown MJ, Sweeting MJ, Brown LC, Powell JT, Thompson SG. Surveillance intervals for small abdominal aortic aneurysms: A meta-analysis. *JAMA*. 2013;309:806–13.
11. Thompson SG, Brown LC, Sweeting MJ, et al. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: Implications for surveillance intervals and their cost-effectiveness. *Health Technol Assess*. 2013;17:1–118.
12. Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the national health service abdominal aortic aneurysm screening programme in England. *Br J Surg*. 2014;101:976–82.
13. Lindholt JS, Juul S, Fasting H, Henneberg EW. Preliminary ten year results from a randomised single centre mass screening trial for abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg*. 2006;32:608–14.
14. Svensjo S, Bjorck M, Wanhainen A. Current prevalence of abdominal aortic aneurysm in 70-year-old women. *Br J Surg*. 2013;100:367–72.
15. Stather PW, Dattani N, Bown MJ, Earnshaw JJ, Lees TA. International variations in AAA screening. *Eur J Vasc Endovasc Surg*. 2013;45:231–4.
16. Wild JB, Stather PW, Biancari F, et al. A multicentre observational study of the outcomes of screening detected sub-aneurysmal aortic dilatation. *Eur J Vasc Endovasc Surg*. 2013;45:128–34.
17. Paraskevas KI, Mikhailidis DP, Andrikopoulos V, Bessias N, Bell SP. Should the size threshold for elective abdominal aortic aneurysm repair be lowered in the endovascular era? Yes. *Angiology*. 2010;61(7):617–9.
18. Powell JT, Thompson SG. Should the frequency of surveillance for small abdominal aortic aneurysms be reduced? *Eur J Vasc Endovasc Surg*. 2013;46:171–2.
19. Earnshaw JJ. Should the frequency of surveillance of a small abdominal aortic aneurysm be reduced? *Eur J Vasc Endovasc Surg*. 2013;46:174.
20. The International AAA Screening Group, Bjorck M, Bown MJ, et al. International update on screening for abdominal aortic aneurysms: Issues and opportunities. *Eur J Vasc Endovasc Surg*. 2015;49:113–5.
21. Davis M, Harris M, Earnshaw JJ. Implementation of the national health service abdominal aortic aneurysm

screening program in england. *J Vasc Surg*. 2013;57(5): 1440-1445.
22. Kim LG. A markov model for long-term cost-effectiveness modelling of screening for abdominal aortic aneurysms: Version 2. http://www.mrc-bsu.cam.ac.uk/wp-content/uploads/A_Markov_model_for_MASS.pdf. Updated 2006. Accessed 9th July, 2016.
23. Kim LG, Thompson SG. Uncertainty and validation of health economic decision models. Health Econ. 2010;19(1): 43-55.
24. Office for National Statistics. Interim life tables, england and wales, 2009-2011. Available from URL: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-274529. Accessed 18 May 2016.
25. Sweeting MJ, Thompson SG, Brown LC, Powell JT, RESCAN collaborators. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *Br J Surg*. 2012;99:655–65.
26. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ*. 2003;12:837–48.
27. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: Selecting the appropriate approach. *J Health Serv Res Policy*. 2004;9:110–8.
28. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: A report of the ISPOR-SMDM modeling good research practices task force-4. *Value Health*. 2012;15:821–7.
29. Screening women for abdominal aortic aneurysms (SWAN). https://www.journalslibrary.nihr.ac.uk/programmes/hta/1417901/#/ Accessed December 10 2016.

APPENDIX D: Paper 4

Paper 4: TOMADO

Sharples L, Glover M, Clutterbuck-James A, Bennett M, Jordan J, Chadwick R, Pittman M, East C, Cameron M, Davies M, Oscroft N. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and Health Technol Assess. 2014;18(67):1. https://doi.org/10.3310/hta18670

Extract presented here is: 'Chapter 4: Long-term cost-effectiveness of oral mandibular devices compared with continuous positive airway pressure and conservative management'

Chapter 4 Long-term cost-effectiveness of oral mandibular devices compared with continuous positive airway pressure and conservative management

Introduction

The results of the within-trial economic analyses based on the TOMADO study data presented in *Chapter 2* showed that that all three of the MADs trialled are cost-effective compared with no treatment for mild to moderate OSAH. The within-trial cost-effectiveness analysis suggests that the SP2, or a similar semi-bespoke device, should be offered as first-line treatment and that dentally fitted bespoke devices should be reserved for those who cannot produce the mould for, or tolerate, a semi-bespoke device. However, there were no statistically significant differences in treatment effects between devices in the base case and results reflect only the observed 4-week follow-up period, comparing each device with no treatment as well as between devices. This chapter presents a cost-effectiveness analysis incorporating long-term effects, to address uncertainties regarding the long-term use of MADs for the treatment of mild to moderate OSAH.

Obstructive sleep apnoea–hypopnoea is a chronic condition and is associated with considerable long-term morbidities, which cannot be fully reflected by a within-trial cost-effectiveness analysis with a short follow-up. For example, large cohort studies have shown that OSAH is associated with hypertension,[133] which will have long-term cardiovascular implications including stroke.[134] The morbidities associated with OSAH are likely to manifest themselves after long-term disease. Excessive daytime sleepiness caused by OSAH also increases the risk of RTAs.[135] These relatively rare events are unlikely to be reflected adequately in short-term trial data.

The long-term and rare events associated with OSAH have survival, QoL and health-care resource use implications, which are important to incorporate in a cost-effectiveness analysis to inform decision-making. While TOMADO's follow-up period was restricted to 4 weeks, partly because of the crossover nature of the trial and the length of follow-up required for gathering data on the primary clinical outcome (AHI), this length of follow-up is common among other studies of interventions to treat OSAH (see *Chapter 3*). To address longer-term cost-effectiveness, several economic models have been developed.[136–142]

Decision-makers also need to be able to compare MADs with other relevant interventions not included in TOMADO. Therefore, an economic model that is able to bring together a range of data sources to chart the long-term morbidities associated with OSAH, as well as symptomatic relief and changes in HRQoL provided by different treatments, is required. The NICE Technology Appraisal 139 defined the potentially suitable treatment options for mild to moderate sleep apnoea as CPAP, MAD or CM.[37] CPAP therapy was recommended in the first instance and oral devices were shown to be cost-effective against CM as an alternative. However, uncertainties remain about the role MADs may play in the treatment of sleep apnoea.

Following a literature search of economic models for OSAH, McDaid *et al.*[8] found a number of key limitations with existing economic evaluations:

- studies did not use the full range of clinical evidence available to estimate the impact of treatment on sleepiness
- a lack of trial-based evidence to compare utility values associated with different treatment options
- limited data on long-term impact of OSAH in terms of cardiovascular risk, RTAs and HRQoL
- the existing evaluations did not examine all the relevant comparators.

To address these limitations, McDaid *et al.*[8] developed a new model to investigate the cost-effectiveness of CPAP compared with MADs and conservative care. To adequately characterise OSAH and its treatment, and ensure that the model was clinically representative, the structure was established from a systematic review used to inform clinical effectiveness, consultation of existing cost-effectiveness literature and opinion of clinical experts involved in the technology assessment process. It made good use of available trial data through a systematic review and meta-analysis of RCTs. The modelling process also followed NICE methodological guidance and used the reference case[37] to increase generalisability.

The perspective, structure, capabilities and treatment options which had been incorporated into the McDaid *et al.*[8] model corresponded to the aims of this evaluation and, therefore, their peer-reviewed model formed the starting point of the long-term economic evaluation. Their conclusion that key uncertainties included the cost-effectiveness of MADs and, hence, the role they should play in the treatment of OSAH, also serves to highlight the importance of the new research in this chapter: 'It remains unclear precisely what type of devices may be effective and in which populations with OSAH. The effectiveness of dental devices compared with CPAP in mild and severe disease populations remains unclear'.[8]

The objectives of the economic analysis presented in this chapter were therefore to update and adapt the York model where necessary to (i) reflect emerging data since the model was built and (ii) focus on the mild/moderate severity patient population. This updated model was then used to assess the cost-effectiveness of MADs, compared with CM and CPAP therapy.

This chapter begins with a summary of the McDaid *et al.*[8] model. It is followed by a description of how parameterisation was completed on the basis of literature searches undertaken to identify potential new sources of data and the incorporation of the TOMADO results into modelling. Results of the analysis of the long-term cost-effectiveness of MADs compared with CPAP and CM for mild/moderate OSAH sufferers are then presented, as incremental cost per QALY. The discussion of these results with the main policy interpretation is left to *Chapter 5*.

The McDaid et al. *model*

McDaid *et al.*[8] developed a state-transition Markov model to assess the long-term cost-effectiveness of CPAP therapy compared with MAD and CM as part of a NICE technology appraisal.[37] The model charted the movement of a hypothetical cohort of 50-year-old men, with characteristics pooled from a meta-analysis of clinical trials of OSAH interventions. Patients were typically overweight (mean BMI = 30 kg/m^2) and had high BP (SBP = 130 mmHg). Baseline EDS, measured by mean ESS score, was 12. Various CPAP devices provided by different manufacturers were treated as one class of intervention. The large numbers of differing MADs used in trials were pooled for an overall treatment effect. CM involved a one-off consultation with a GP, with some level of lifestyle advice on how to reduce or cope with symptoms better. Outcomes were summarised as an incremental cost per QALY for each intervention. The model structure is explained briefly below.

Given the chronic nature of OSAH, the McDaid *et al.*[8] model adopted a lifetime horizon and incorporated the possibility of CVEs, strokes and involvement in RTAs, as well as accounting for symptomatic effects of OSAH on QoL. Patients started in an OSAH state and were able to move into a number of different health states [OSAH post coronary heart disease (CHD), OSAH post stroke and death], reflecting morbidities linked to long-term OSAH suffering. The model ran on a yearly cycle to chart a hypothetical cohort of 10,000 patients over time.

Figure 30 provides a diagrammatic representation of the model. Elliptical boxes represent health states and square boxes represent events. Arrows show the direction of transitions between health states and the occurrence of events. All members of the cohort started in the OSAH state and could stay in that state, unless a transition occurred, until death. They could move into the post-CHD state if they experienced an acute CVE and survived. This state allowed for the increased morbidity and mortality associated with having had a first CHD event. If they did not survive, they moved to the absorbing death state. If they did

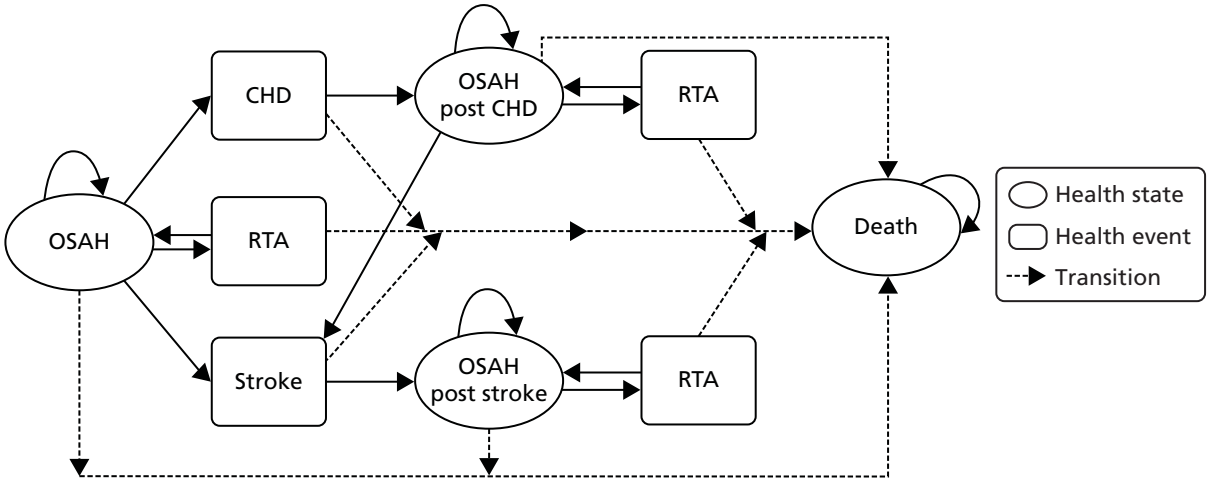


FIGURE 30 Long-term model structure developed by McDaid *et al.*[8]

survive, they could remain in this post-CHD state until death, or experience a RTA (fatal or non-fatal) or suffer a stroke. If they survived a RTA, they remained in the same health state post event. If they survived a stroke, they moved to the OSAH post-stroke state, where they were again able to remain until death or experience a RTA. They were not able to move back to a CHD state once they had suffered a stroke. Patients who had a disabling stroke were assumed to no longer be able to drive and, hence, a proportion of those in the post-stroke state were not able to have a RTA event.

Patients could suffer a stroke while in the initial OSAH state, in which case, if they survived, they would move to the post-stroke health state. Here they would be subject to the increased risk of mortality and morbidity following the first event. Provided the stroke was not disabling they could experience a RTA (fatal or non-fatal). Patients in the initial OSAH state may at some point have experienced a RTA and, provided it was not fatal, would stay in the OSAH state until another transition or death.

Movements between states were determined by a set of transition probabilities, derived from various sources. In the base case, transitions that relate to CVEs and risk of stroke were informed by the Framingham risk equation, utilising information on baseline characteristics of an OSAH population to calculate the probability of a CVE (*Table 40*). Differences in SBP observed under the treatment options (from a meta-analysis of RCTs) were used in the Framingham equation to differentiate the risk of CVEs and strokes under each intervention. The equation is based on Weibull models, meaning that predicted risk is non-linear with respect to each risk factor. McDaid *et al.*[8] tested whether or not use of mean BPs would

TABLE 40 Model cohort characteristics for use in the Framingham equation

Parameter	Mean	Source
Age (years)	51	TOMADO mean
SBP	130	TOMADO mean
Smoking (0 = no; 1 = yes)	0	Assumption (TOMADO 25% smokers)
Total cholesterol (mg/dl)	224	Coughlin *et al.*[143]
HDL cholesterol (mg/dl)	43	Coughlin *et al.*[143]
Diabetes (0 = no; 1 = yes)	0	Assumption (TOMADO 7% diabetic)
ECG-LVH (0 = no; 1 = yes)	0	Assumption
Baseline ESS score	11.9	TOMADO mean

ECG, electrocardiogram; HDL, high-density lipoprotein; LVH, left ventricular hypertrophy.

bias the results using a set of individual patient data. From the equation, the risk of CVEs and stroke were predicted using BP for each patient, and the mean taken. This was compared with risk calculated based on the mean of group BPs. The risk calculated by the two different methods was the same to two decimal places and, so, use of aggregate-level data did not significantly bias results. The equation was used to calculate the 4-year probability of an event, with a piece-wise exponential used to convert this into a yearly probability to correspond to the cycle length.

Long-term observational studies were consulted for estimates of the increased risk of mortality following events relating to stroke and CHD once an initial event had occurred.[144,145] The underlying risk of RTAs (fatal and non-fatal) was estimated from Department of Transport[146] data and was adjusted based on the OR of RTAs given treatment with CPAP compared with no treatment, taken from an updated meta-analysis by Ayas *et al.*[136] Given a lack of data on the likelihood of a RTA when using MADs, the ratio of ESS scores for MAD treatment compared with CM was applied to the OR for RTAs of CPAP compared with CM. Symptomatic relief provided by different interventions was accounted for using evidence from a meta-analysis of ESS scores, which were mapped to a QoL scale, in the absence of good HRQoL data. Regression techniques were used to estimate an algorithm for expressing utility changes, as measured by EQ-5D-3L and SF-6D pre-scored preference questionnaires to changes in ESS score. Utilities and costs were assigned to each of the health states and differed depending on the intervention being received. Each health event had an associated utility loss and acute cost attached to the event.

Costs of interventions were estimated in 2005 prices (£), incorporating the cost of devices and any on-going resource usage associated with maintenance and replacement, including equipment, staff time and overheads. CPAP device costs were acquired from McDaid *et al.*[8] Estimation of resource use during the titration process was taken from a manufacturer's submission to NICE, which included data elicited from a group of clinicians regarding proportion titrated by different methods in clinical practice to ascertain appropriate costs. The machine was assumed to have a lifespan of 7 years (clinical opinion) and masks replaced annually. It was assumed that the MADs being used was a Thornton Adjustable Positioner® (Airway Management Inc., Dallas, TX, USA), commonly in use at the time and this was costed according to NHS Dental contract costs, given the lack of an appropriate NHS cost of the device. The lifespan of a MAD was assumed to be 2 years (clinical opinion) compared with 12–18 months in the TOMADO study. Unit costs for NHS resource use (sleep specialist consultations, nurse appointment and GP consultations) were taken from nationally available NHS reference costs, as well as unit costs published by the Personal Social Services Research Unit (PSSRU).[58,147] Published sources were consulted for estimates of the cost of other morbidities (CHD, stroke and RTAs) associated with OSAH. Two economic evaluations which had estimated costs of an acute CHD (and on-going treatment costs of chronic conditions) and stroke events in a NHS setting were used.[148,149] RTA costs were taken from UK Department of Transport estimates.[146] Cost and effects were discounted at 3.5% per annum.

The modelling was implemented in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) and results presented as ICERs representing the long-term mean cost per QALY gained for the different interventions. Uncertainty was explored using probabilistic techniques, by attaching distributions to input parameters and randomly sampling from them, performing 10,000 iterations to produce estimates of the distributions of the outcome. This uncertainty was summarised using CEACs, showing the likelihood that any given device is cost-effective at a given WTP threshold.

Results from McDaid *et al.*[8] indicated a 78% probability that CPAP was cost-effective for the hypothetical cohort at a threshold of £20,000 per QALY. At this WTP, MADs and CM had a probability of being cost-effective of 21% and 1%, respectively. Sensitivity analysis suggested that CPAP had the highest probability of being cost-effective over a wide range of WTP thresholds, even for mild and moderate subgroups, though the probability of MADs being cost-effective increased for milder subgroups.

Updating model parameter values

For this cost-effectiveness analysis, the parameters used to populate the economic model were revisited, to update where necessary and possible. Treatment effects were restricted to a mild/moderate severity group of OSAH sufferers and taken from the meta-analyses presented in *Chapter 3*, which incorporated both TOMADO and other RCT data. Within-trial effects were used in a sensitivity analysis to investigate potential between device differences in long-term cost-effectiveness. Other data from TOMADO used in the model included costs and HRQoL. The remaining data for the economic model were produced following replication of searches first performed by McDaid *et al.*[8] on cardiovascular risk and sleep apnoea, HRQoL data, and RTA risk and sleep apnoea. A new review on compliance of CPAP and MADs was also conducted. The decision about whether or not new evidence was chosen in preference to that already parameterising the model was based on the following criteria:

- evidence was specific to a mild to moderate OSAH population
- estimates were UK specific or more relevant to the NHS
- data were more robust (based on characteristics such as sample size and study design)
- evidence was contemporary compared with previous estimates or
- new evidence facilitated improved modelling (for instance longer-term data or enabling structural improvements) of OSAH and its treatment.

Cardiovascular risk and obstructive sleep apnoea–hypopnoea syndrome

McDaid *et al.*[8] recognised CVEs as a major source of morbidity associated with OSAH and modelled accordingly. Based on literature searches, the evidence established a link between OSAH and CVD, the strongest with regards to OSAH being a risk factor in hypertension,[150,152] though there remained some doubt about whether or not it is an independent risk factor. For this reason, and given a lack of data on long-term outcomes for treatment of OSAH, CVEs were linked to OSAH using a risk score which accounts for the increased risk from raised BP.

In order to account for uncertainties around OSAH and cardiovascular risk, assess the current understanding of the link between OSAH and CVD, and allow for any long-term evaluation of interventions, the literature search of CVD and its role in OSAH was updated. Although some of the RCTs identified by the systematic review in *Chapter 3* had investigated longer-term CVD outcomes under treatment, the majority did not and instead focused on intermediate outcomes, mainly BP. Follow-up was often not sufficiently long to capture these rare events.

Literature search

A search of MEDLINE for 2007–2013 to find articles that referenced OSAH and CVD used a subset of terms that could be encompassed into CVD (e.g. stroke, heart disease, hypertension) which was very similar to that performed by McDaid *et al.*[8] (see details in *Appendix 14*). The original search had also looked for RTA literature, but this was left to an additional search. The search yielded over 500 papers, which were screened by title and abstract. The focus was on identifying new analyses of primary data, including observational studies not identified as part of the systematic review of *Chapter 3* and previous reviews. The majority were excluded as they were not related to OSAH, and 82 were shortlisted, of which 24 were examined in more detail. The 57 excluded were guidelines, commentaries, editorials, letters or case reviews ($n = 18$); duplicates or duplicating clinical trial data already identified in the systematic review (see *Chapter 3*) ($n = 2$) [e.g. referring to a different patient population (e.g. focused on central apnoeas or a younger population) ($n = 16$)]; did not consider the association between OSAH and CVD risk ($n = 10$); were not in the English language; or had only abstracts available ($n = 11$). Owing to the heterogeneity between studies in methodology and markers of hypertension used, a narrative review is provided rather than a formal meta-analysis.

Several studies explored the link between OSAH and CVD. Two studies showed the high prevalence of cerebrovascular lesions[153] and hypertension[154] among an OSAH population. In the former,[153] the prevalence of silent lacunar infarction among 192 patients with moderate and severe OSAH

(AHI ≥ 15 events/hour) was higher than among the controls and the patients with mild OSAH ($p < 0.0001$). In a population of 125 hypertension sufferers, OSAH was present in 64%, a much higher prevalence than in the general population.[154] A small case–control study ($n = 50$) found that nearly 60% of patients who had had a stroke and ischaemic attacks displayed OSAH.[155] In a case–control study (63 cases and 63 matched controls), patients with resistant hypertension (inclusion criteria: BP > 140/90 mmHg, using at least three BP-lowering drugs, including a diuretic), 45 of the case subjects were found to be OSAH sufferers compared with 24 of the controls ($p < 0.001$).[156] Logistic regression gave those with OSAH an OR for suffering from resistant hypertension of 4.8 (95% CI 2.0 to 11.7). A case-matched study of 227 OSAH patients used multiple variable regression to estimate an OR for coronary heart failure of 5.47 (95% CI 1.06 to 28.31) for OSAH sufferers compared with controls.[157]

Several articles analysed data from large cohort studies, with mixed results indicating OSAH as an independent risk factor for hypertension:

- Young *et al.*,[133] in a subset of data from the Wisconsin Sleep Cohort ($n = 1549$), found an OR for 4-year incidence of hypertension (defined as BP > 140/90 mmHg or treatment with antihypertensives) of 2.0 (95% CI 1.2 to 3.2) for patients with an AHI of 5–15 events/hour compared with patients with an AHI < 5 at baseline; patients with an AHI > 15 had an OR for 4-year incidence of hypertension of 2.9 (95% CI 1.5 to 5.6) compared with patients with an AHI < 5 at baseline.
- Marin *et al.*[158] looked at a cohort of control subjects (AHI < 5 events/hour) and OSAH sufferers ($n = 1889$) treated with CPAP therapy. They estimated an adjusted HR for incident hypertension compared with controls which was greater among patients with untreated OSAH; among those ineligible for CPAP therapy, HR was 1.33 (95% CI 1.01 to 1.75), compared with 1.96 (95% CI 1.44 to 2.66) among those who declined CPAP therapy and 1.78 (95% CI 1.23 to 2.58) among those non-adherent to CPAP therapy. All displayed higher rates of hypertension than control subjects.
- O'Connor *et al.*[159] using data from the Sleep Heart Health Study ($n = 2470$ men) after a mean of 2 years of follow-up and based on the same definition of hypertension, observed an OR (adjusted for age, sex, race and time since baseline) of 2.19 (95% CI 1.39 to 3.44) for people with an AHI of > 30 events/hour compared with an AHI of 0.0–4.9 events/hour, though this relationship became weaker (and not significant) for lower AHI. When adjusted for further baseline characteristics (BMI, waist-to-hip ratio and neck circumference) the OR was 1.50 (95% CI 0.91 to 2.46) suggesting a moderate but not significant association, which was again further weakened for lower AHI.
- Kapur *et al.*[160] used the same dataset and demonstrated that the relationship is stronger if patients are stratified by AHI and sleepiness. They estimated an adjusted OR of 3.04 (95% CI 1.33 to 6.04) for an AHI > 30 and experiencing frequent sleepiness (≥ 5 days).
- Using the same definition of hypertension (based on BP or taking hypertensive medication), the Vitoria Sleep Cohort[161] of 1180 patients showed similar results. The crude OR suggested an association, with respiratory disturbance index (RDI) of > 14 compared with 0.0–2.9 giving an OR of 2.61 (95% CI 1.75 to 3.89). An OR greater than 1 held for lower strata of RDI, which were all significant. However, when adjusted for age, sex, BMI, neck circumference, alcohol, coffee and tobacco consumption, and fitness level the OR for RDI > 14 compared with an RDI of 0.0–2.9 was 0.98 (95% CI 0.62 to 1.57), which suggests obstructive sleep apnoea (OSA) is not an independent risk factor.

Other data from the Sleep Heart Health Study ($n = 5422$) suggest that OSAH is associated with a higher chance of suffering a stroke (OR 2.86, 95% CI 1.10 to 7.39, at an AHI of > 19 events/hour).[4] The point estimate of the OR was similar in lower severity OSAH, but the difference was not statistically significant. Martínez-Garcia *et al.*[162] undertook a prospective observational study offering CPAP to OSAH patients, with 7 years' mean follow-up ($n = 223$) of non-fatal CVEs. For a group of patients with an AHI > 20 who had not been able to tolerate CPAP, they estimated a HR, using Cox-adjusted proportional regression, of 2.87 (95% CI 1.11 to 7.71).

Several of the articles ($n = 9$) were review papers combining existing prospective evidence on the association between hypertension, CVD (including stroke), mortality and OSAH.

Several reviews examined the mechanisms involved in OSAH's role in hypertension.

In a 2009 review, Bradley and Floras[5] state: 'Data from animal models, epidemiological studies, and RCTs provide strong evidence that OSAH can cause hypertension, and that its treatment can lower BP. Indeed, OSAH might well be the commonest treatable cause of secondary hypertension.' The same authors were involved in a subsequent review in which Kasai *et al.*[163] noted the higher prevalence of OSAH among a CVD population (47–83%). They suggest that repetitive apnoeas expose the heart and circulatory system to 'noxious stimuli' which can lead to CVD through OSAH's causal role in negative intrathoracic pressure, autonomic dysregulation, oxidative stress, inflammation, endothelial dysfunction, platelet activation and hypercoagulability. Although no quantitative synthesis of data was undertaken, Kasai *et al.*[163] asserted that 'data from epidemiological studies and randomised clinical trials strongly suggest that OSA is a common and treatable risk factor for development of hypertension, heart failure, arrhythmias, and stroke, especially in men'. However, they also proposed that the relationship may be bidirectional. Kato *et al.*[164] also conclude that the pool of evidence relating OSAH to CVD is growing, and state that this is strongest in relation to the role of OSAH in hypertension. Monahan and Redline[165] corroborate assertions around improved understanding of pathophysiological basis of the association of OSA and CVD and note the 'modest improvements in BP associated with continuous positive airway pressure (CPAP) use'.

Two reviews note that BP is lowered by treatment of OSAH. Calhoun[166] explores the mechanism of OSA-induced hypertension and presents results of four meta-analyses suggesting that BP is lowered by CPAP treatment [SBP lowered by 1.38 mmHg (not significant), 2.46 mmHg, 1.64 mmHg and 0.95 mmHg (not significant)] and data included in Monahan and Redline[166] corroborate this. No cohort studies that show long-term treatment effects (with estimates of ORs or relative risks) for interventions used to treat OSAH were identified.

Several reviews also highlighted the role of OSAH in stroke.

Loke *et al.*[134] conducted a meta-analysis which included nine prospective studies ($n = 8400$) investigating OSAH and CVD outcomes and suggested an association between OSAH and strokes (OR 2.24, 95% CI 1.57 to 3.19) and heart disease (OR 1.56, 95% CI 0.83 to 2.91), though the relationship was not statistically significant for the latter. Wallace *et al.*[167] conducted a qualitative review of sleep-related disorders and stroke. The authors comment on the established association between OSAH and stroke, citing evidence from the Sleep Heart Health Study and the Wisconsin Sleep Cohort referred to earlier, and state the case for screening stroke patients for OSAH. In another review, Dyken and Im[168] conclude that OSAH is independently associated with a range of stroke factors but note that, while there is some evidence that treatment can reduce BP, there is a lack of definitive RCT data on overall stroke risk. Portela *et al.*[169] and Caples[170] echo the findings of both these reviews.

However, recognition of a lack of good trial data was a recurrent theme. Monahan and Redline[165] allude to the need for well-powered clinical trials investigating long-term CVD outcomes in OSAH under treatment. Kohli *et al.*[171] and Parati *et al.*[172] make similar conclusions regarding the gaps in current evidence.

While the role of OSAH in CVEs is still somewhat unclear, new evidence does suggest an association. However, there is still a lack of good-quality evidence on the long-term cardiovascular and stroke outcomes of treatment of OSAH, for patients using both CPAP and MADs. There is greater understanding since McDaid *et al.*[8] addressed the literature, of the potential causal factors relating to OSAH and CVD and stroke,[164,165] but they are probably multifactorial and may be bidirectional.[163]

As McDaid *et al.*[8] found, evidence still seems to be strongest in supporting the role of OSAH in hypertension. Analysis of data from large cohort studies (the Wisconsin Sleep Cohort)[133] showed an association, especially among men, but there remains conflicting evidence (The Sleep Heart Health Study;[159] Vitoria Sleep Study[161]). Based on these findings and the BP data found in randomised trials,

the use of the Framingham risk equation was not modified on the basis of data published since the McDaid *et al.* modelling exercise. The characterisation of risk through an algorithm such as the Framingham equation, which uses differences in BP to differentiate CVE risk between baseline and post intervention, seems appropriate given the lack of good data on long-term outcomes. Baseline risk is defined by characteristics taken from TOMADO and a study investigating the role of OSA and metabolic syndrome by Coughlin *et al.*[143] Other cardiovascular inputs to the model are given in *Table 41*.

While the Framingham equation was used in the base case, an additional source of the relative risk associated with a reduction in SBP was identified. Lewington *et al.*[174] pooled data from 61 cohort studies to estimate the relationship between BP and vascular mortality. Adjusting for regression dilution, at ages 60–69 years the relative risk of a stroke for a 20 mmHg reduction in SBP is 0.43 and the relative risk of CHD is 0.54. Given the linear relationship, a proportional change for a 1 mmHg reduction was taken. This analysis also suggests that the reduction in risk is proportional, independent of pre-treatment BP. The baseline risk from the Framingham equation was taken. The proportion of disabling strokes was taken from a large RCT of over 6000 patients comparing interventions for secondary prevention of vascular events.

Road traffic accident risk

To incorporate the change in risk of RTAs following treatment for OSAH, McDaid *et al.*[8] updated a meta-analysis first undertaken by Ayas *et al.*[136] with one additional study by Barbé *et al.*[175] with the eight studies in the Ayas *et al.*[136] review. All of these studies had before-and-after designs, based on actual RTA events pre- and post-CPAP therapy. Barbé *et al.*[175] collected 2 years of collision information retrospectively from participants prior to the study and then prospectively recorded events for 2 years while using CPAP. This study reported a relative risk, but gave event numbers which were used to calculate an OR compatible with the Ayas *et al.*[136] data. Results from the nine studies were pooled to give an OR of 0.168 (95% CI 0.100 to 0.230) after treatment with CPAP. This suggests that the odds of a RTA are reduced by nearly six times when CPAP treatment is initiated. While this effect size is quite large, the underlying rate of a RTA[176] was extremely low (non-fatal: male = 0.0089 per year, female = 0.0082 per year; fatal: male = 0.00014 per year, female = 0.00006 per year).

The rates of RTAs in the model were updated using data derived from the National Travel Survey for 2010[177] and UK Data Archive data from 2010 on RTAs,[178] which presented equivalent contemporary data to those used by McDaid *et al.*[8] The risk was calculated based on the number of UK driving licences held and the numbers of fatal traffic accidents and traffic accidents involving serious and slight injury for 2010. These rates are given in *Table 42*.

TABLE 41 Coronary heart disease and stroke parameters

Parameter	Mean	SD	Source
Relative risk of death following CHD	3.2	0.30	Rosengren *et al.*[144]
Relative risk of death following stroke	2.3	0.18	Dennis *et al.*[145]
Proportion of strokes that are disabling	0.309	–	Diener *et al.*[173]

TABLE 42 Underlying risk of RTAs

Parameter	Mean	SD	Source
Rate of non-fatal RTAs for males	0.0062	pop^n	Department of Transport[146]
Rate of fatal RTAs for males	7.11×10^5	pop^n	Department of Transport[146]
Rate of non-fatal RTAs for females	0.0053	pop^n	Department of Transport[146]
Rate of fatal RTAs for females	2.91×10^5	pop^n	Department of Transport[146]

The search used by McDaid *et al.*[8] was rerun to identify new studies conducted between 2007 and 2013 relating to OSAH and the risk of RTAs.

Literature search

The search (see terms used in *Appendix 14*) identified 32 articles, which were screened for relevance. Nineteen were excluded on the basis that they were commentaries or editorials ($n = 3$); duplicates ($n = 1$), referred to the wrong patient population (e.g. non-OSAH patients, elderly population) ($n = 5$); did not consider RTA risk ($n = 6$); were not in the English language; or only had abstracts available ($n = 4$). Of the 13 studies reviewed in greater detail only two related to observed RTA risk post treatment.[179,180] These two articles were meta-analyses of RTA risk post OSAH treatment. One additional study considered simulated driver performance before and after CPAP treatment.[181] The other nine included clinical effectiveness and cost-effectiveness studies and case–control studies comparing OSAH risk with healthy populations.

The two new meta-analyses pooling data on the impact of CPAP on RTAs were:

- Tregear *et al.*[180] analysed nine studies, including one additional study by Scharf *et al.*[182] that did not appear in the Ayas *et al.*[136] and McDaid *et al.*[8] meta-analyses. However, the Tregear *et al.*[180] analysis also omitted one study by Suratt and Findley[183] that Ayas *et al.*[136] and subsequently McDaid *et al.* had included. The Suratt and Findley[183] article is available only in abstract form and may have been excluded by Tregear *et al.*[180] given their criteria that all studies must be published in full. It is not clear why the study by Scharf *et al.*[182] was not included in the Ayas *et al.*[136] review, which McDaid *et al.*[8] subsequently updated. Tregear *et al.*[180] estimated an OR of 0.278 (95% CI 0.220 to 0.350) for the risk of a RTA post CPAP treatment compared with pre-intervention. This is higher than, but comparable to, the OR of 0.168 estimated by McDaid *et al.*[8]
- Antonopoulos *et al.*[179] performed an analysis of real accidents, accident near misses and simulated driving performance. Ten studies of real accidents (including the Suratt and Findley[183] data) were included. As in the review by Ayas *et al.*,[136] the Scharf *et al.*[182] study was absent, but this review did include another study by Minemura[184] that was not in the McDaid *et al.*[8] or Tregear *et al.*[180] analyses. While the study by Minemura[184] may have been excluded by Tregear *et al.*[180] because of their inclusion criterion that studies should involve more than 20 patients, the reason for omission from the Ayas *et al.*[136] review is unknown. An OR of post-CPAP compared with pre-CPAP RTA risk of 0.21 (95% CI 0.12 to 0.35) was estimated and pooled data on driving simulator performance showed a significant improvement in performance post treatment.

An additional study, by Hoekema *et al.*,[181] based on a prospective simulator-based investigation of driving performance of 20 OSAH patients and 16 controls, was also found. OSAH patient simulator performance was compared with the control group before and after 8 weeks of CPAP ($n = 10$) and MAD ($n = 10$) treatment. Patients randomised to each group were subject to 25 minutes of driving simulation and lapses of attention were observed. The results suggested significant differences in performance post treatment, similar for both CPAP and MADs.

Given the difficulty in ascertaining the reason for inclusion of studies and the effect it leads to in differences of ORs pooled by the two new meta-analyses and the McDaid *et al.*[8] analysis, the OR of experiencing a RTA of 0.17 from McDaid *et al.*[8] was retained. The two newly identified estimates of the reduction in RTA risk post treatment are of similar magnitude, but the Tregear *et al.*[180] estimate was used in a scenario analysis, as it suggested the smallest effect size. Given that this estimate is specific to CPAP, the approach to the comparison of MADs with CM followed the method of McDaid *et al.*[8] That is, a multiplier based on the relative treatment effects on ESS score of CPAP versus CM and MADs versus CM was applied to the OR of RTA for CPAP versus CM. These rates are presented along with other treatment effects in *Table 42*.

Health-related quality of life

During their systematic review, McDaid *et al.*[8] highlighted the paucity of data regarding HRQoL and OSAH. To characterise cost per QALY using the NICE reference case, utility scores are needed for each treatment.

As these were lacking, and a large number ($n = 27$) of the trials in the systematic review of treatment effects had reported ESS scores, McDaid *et al.*[8] used the surrogate end point of ESS score as a proxy for differences in utility. Three sets of individual patient-level data (two measuring ESS and SF-36 profile in the same patients and one that measured ESS, SF-36 profile and EQ-5D-3L data in the same set of patients) were used to map ESS scores to EQ-5D-3L and SF-6D values (based on tariffs published by Brazier *et al.*[63] and Dolan[61]) using regression analyses. The results of this process indicated that a unit fall in ESS score is associated with an increase in utility, based on a SF-6D ($n = 294$) value of 0.0095 (95% CI 0.0070 to 0.0123) and based on an EQ-5D-3L ($n = 94$) value of 0.0097 (95% CI 0.0019 to 0.0175).

The systematic review presented in *Chapter 3* highlights the remaining dearth of RCT data on OSAH and HRQoL. In trials that did include some measurement of QoL, it was predominantly limited to disease-specific measures (SAQLI and FOSQ). However, one study did use generic instruments to measure HRQoL. In a double-blind randomised trial of 102 men who received a real or sham CPAP device, Siccoli *et al.*[18] used the SF-36 and SF-12 4 weeks after treatment to measure impact of CPAP therapy on HRQoL. This population was defined as having moderate/severe OSAH. In the intervention group, scores on several domains of the SF-36 (Emotional Well-being, Vitality, Role Emotional and Social Function) were significantly higher than those in the sham group. Using the SF-12, the mean PCS difference was 58.8 compared with 72.4 and the mean MCS was 63.5 compared with 77.9, both differences being 'significant'. However, a utility score based on these short-form surveys was not presented.

While TOMADO included the EQ-5D-3L and SF-36, these data were relatively short term and specific only to MAD. Therefore, further searches were undertaken to identify other potential sources of HRQoL utility data from generic instruments, for use in the modelling.

Literature search

A search first performed by McDaid *et al.*[8] was replicated for 2007–13, using MEDLINE, to identify data not included in the systematic review reported in *Chapter 3*, i.e. including observational trials that might offer a robust data source.

The search yielded over 700 potentially relevant articles, which were screened by title and abstract for relevance. The aim was to identify studies of OSAH which included a treatment (either CPAP or MADs) and measured QoL using the EQ-5D-3L or SF-36/SF-12 pre-scored preference questionnaires. Seventy-one papers were examined in greater detail (see list in *Appendix 15*, along with the search terms).

Of the 72 papers examined further, only two captured generic QoL data. A prospective study by Tsara *et al.*[185] reported SF-36 profiles for 135 patients (120 with severe and 15 with mild/moderate OSAH based on AHI) before and after CPAP therapy in a sleep unit at a general hospital in Greece. These data suggested improvements in QoL post CPAP treatment, though this was not expressed as a utility score. Improvements for men were observed in all domains except Pain (Physical Role, Physical Function, Emotional Well-being and Vitality: $p < 0.01$; General Health, Role Emotional and Social Function: $p < 0.05$), with the greatest change in General Health. Women displayed a significant improvement only in Role Physical ($p < 0.01$). Antic *et al.*[186] collected data as part of a randomised trial of nurse-led care for moderate to severe OSAH patients. One hundred and thirty-five OSA patients were included, with SF-36 measurement 3 months after treatment with CPAP in three sleep centres in Australia. SF-36 domains were not presented, but the authors reported that the vitality component was significantly correlated with objective adherence.

Neither of the additional studies[185,186] considered MADs and the focus was in a more severe disease group than TOMADO is primarily focused on. The study by Siccoli *et al.*[18] does offer some robust trial data regarding CPAP treatment effects that could have been converted to SF-6D but, again, these are in a group of patients with moderate to severe disease. TOMADO collected data for a mild to moderate group using MADs, which suggested there may be some improvements in HRQoL after treatment, but these results were not significant for generic instruments.

Bearing in mind these limitations and the desire to utilise the synthesised systematic review of treatment effects, the clinical end point of ESS was again mapped to utility. As TOMADO provided more data points than had been available to McDaid *et al.*,[8] these data were used to estimate a relationship, mapping observed ESS scores to utility measures. The resulting algorithm then converted ESS score treatment differences into post-treatment utility changes.

Mapping Epworth Sleepiness Scale score to European Quality of Life-5 Dimensions three-Level version and Short Form questionnaire-6 Dimensions

TOMADO presented a large dataset of both SF-36 and EQ-5D-3L data for people with mild to moderate OSAH. Given repeated measurements, it yielded 402 data points of ESS score and SF-6D and 404 data points of ESS score and EQ-5D-3L that could be used in a regression-based mapping exercise to estimate an algorithm mapping ESS to utility scores. The algorithms for SF-6D and EQ-5D-3L were estimated using a linear mixed-effects regression model. The ESS score was an explanatory variable; a dummy variable was used to control for differences in baseline utility and participants were included as a random effect. These models rely on an assumption that the residuals are Normally distributed, though this may not always hold.[187] The models are shown in *Table 43* for SF-6D and *Table 44* for EQ-5D-3L.

Figure 31 shows that the residuals appear to be reasonably close to normality for SF-6D, but less so for the EQ-5D-3L. This is consistent with our a priori knowledge of the discrete nature of the EQ-5D-3L, the ceiling effect often observed in relatively healthy groups of patients[188] and the findings of the McDaid *et al.*[8] mapping exercise. Other studies of utility indices derived from EQ-5D-3L in OSAH sufferers confirm this phenomenon and suggest that SF-6D may display a distribution closer to normality.[189]

The results of this regression analysis indicate that a 1-unit decrease in the ESS is associated with a 0.0061 ($p < 0.001$) rise in utility based on EQ-5D-3L and a 0.0067 ($p < 0.001$) rise in utility based on the SF-6D instrument. For probabilistic sensitivity analysis the estimated variance matrix from the linear mixed models was used when sampling from the parameter distributions. The baseline utility of the population in the economic model was estimated based on the mean baseline ESS score of patients in TOMADO. The coefficients in the mapping equations estimated from the TOMADO data were similar to, but slightly lower than, those estimated by McDaid *et al.*[8] This should be expected as the population of patients recruited to TOMADO had mild to moderate OSAH and so represented a subsection of the range of disease severity.

Treatment effects of the use of MADs and CPAP from the meta-analyses of ESS scores in *Chapter 3* were converted into utility increments using the algorithm. The baseline utility was estimated based on the mean ESS score of the trial participants in the TOMADO. The utilities used in the model are shown in *Table 45*.

TABLE 43 Mixed-effects model for mapping ESS scores and utility based on SF-6D ($n = 402$)

Variable	Coefficient	SE	*p*-value	95% CI
ESS	–0.0067	0.0011	0.000	–0.0087 to –0.0046
Baseline	–0.0020	0.0079	0.799	–0.0175 to 0.0134
Constant	0.7529	0.0116	0.000	0.7302 to 0.7756

TABLE 44 Mixed-effects model for mapping ESS scores and utility based on EQ-5D-3L ($n = 404$)

Variable	Coefficient	SE	*p*-value	95% CI
ESS	–0.0061	0.0020	0.003	–0.0101 to –0.0020
Baseline	0.0139	0.0145	0.340	–0.0146 to 0.0423
Constant	0.9094	0.0220	0.000	0.8664 to 0.9525

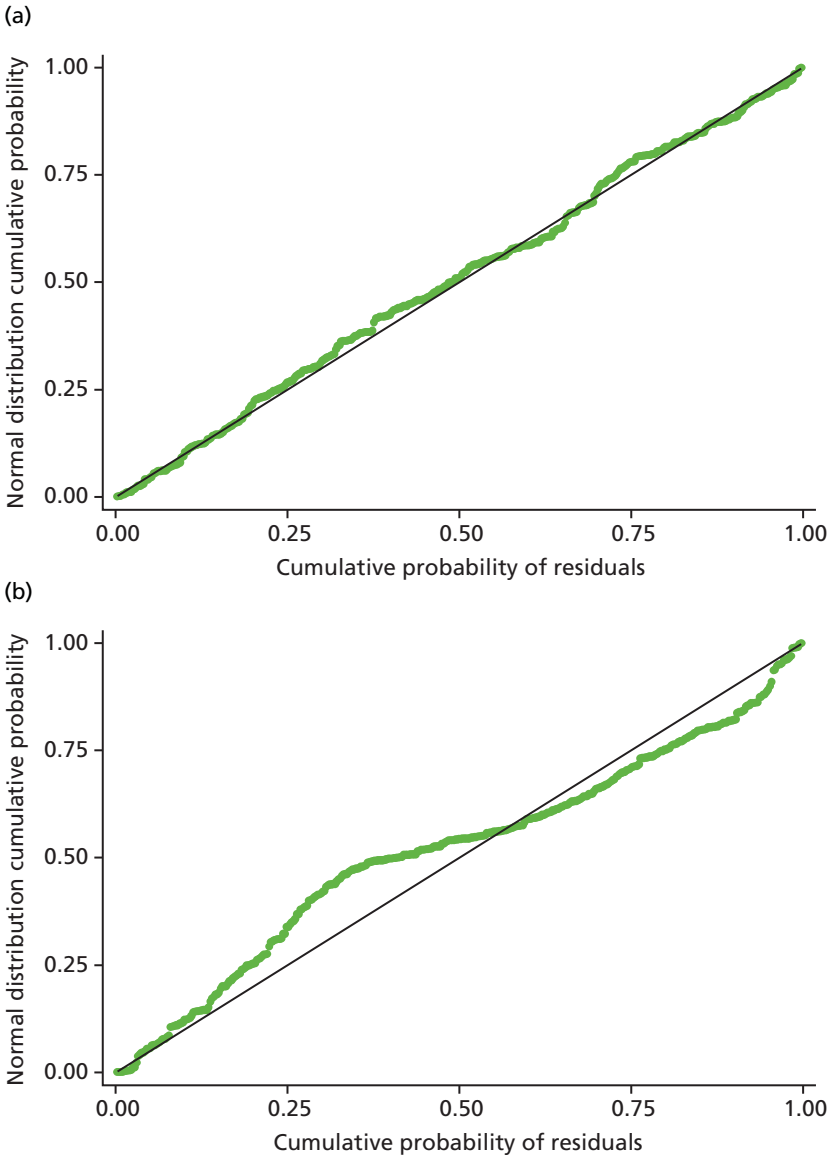


FIGURE 31 Residuals from linear model mapping ESS to (a) SF-6D; and (b) EQ-5D-3L utility scores.

TABLE 45 Utilities for CVEs and RTAs

Utility	Mean	SD	Source
OSAH untreated (baseline)	Baseline ESS score × −0.006 + 0.91	–	TOMADO EQ-5D-3L mapping algorithm
OSAH treated with MAD	$\Delta ESS_{MAD\text{-}CM} \times -0.006$	–	TOMADO EQ-5D-3L mapping algorithm
OSAH treated with CPAP	$\Delta ESS_{CPAP\text{-}CM} \times -0.006$	–	TOMADO EQ-5D-3L mapping algorithm
Stroke (decrement)	−0.0524	0.0002	Sullivan and Gushchyan[190]
CHD (decrement)	−0.0635	0.0001	Sullivan and Gushchyan[190]
RTA	0.6200	0.2700	Currie *et al.*[191]
Age (decrement per year)	−0.0007	0.0000	Sullivan and Gushchyan[190]

McDaid *et al.*[8] relied on data from a study conducted by Sullivan and Ghushcyan,[190] which used EQ-5D-3L data from a panel of 38,678 patients to estimate decrements associated with a range of chronic diseases. The utility associated with a RTA was based on EQ-5D-3L data from a data repository 6 weeks after an inpatient episode for injuries sustained from a RTA in the UK. No additional robust sources of utility data were identified and these values were retained.

Compliance

McDaid *et al.*[8] used a study by McArdle *et al.*[45] of long-term (median follow-up = 1.8 years) CPAP use in Scotland ($n = 1155$) to inform compliance in the model. This prospective observational study collected data on patients offered CPAP therapy. The mean ESS score of patients starting CPAP at baseline was 12 and AHI was 30 events/hour. Patients who refused CPAP therapy had a lower mean AHI of 22, though this was not shown to be a significant predictor of CPAP acceptance. Continued CPAP usage was significantly associated with AHI, with a HR estimate (relative risk of stopping CPAP) using Cox proportional regression of 2.48 (95% CI 1.79 to 3.40) for AHI < 15 relative to AHI ≥ 15. The study also reported a HR for stopping CPAP of 1.92 (95% CI 1.41 to 2.61) for an ESS score < 10 relative to an ESS score > 10. A Kaplan–Meier curve of CPAP use over 5 years was used to calculate yearly probabilities of patients stopping CPAP. The proportion still using CPAP was 0.84 at year 1, 0.74 at year 2, 0.73 at year 3 and 0.68 at year 4. After 4 years, a plateau was observed and, so, it was assumed that all patients who had not stopped using CPAP would continue to use the device indefinitely. In the absence of equivalent data for MADs, McDaid *et al.*[8] assumed compliance was equal to that of CPAP.

A search was conducted to identify new compliance data for both MADs and CPAP.

Literature search

The search of MEDLINE yielded 111 articles that were screened by title and abstract. The terms used are in *Appendix 14* and selection focused on long-term estimates. Studies were considered relevant if they included the use of MADs or CPAP for treatment of OSA and had at least 1 year mean follow-up, indicating a measure of compliance over time. Studies were limited to those with at least 50 patients. Thirty-eight were reviewed in more detail. Of these, many did not have at least a year of follow-up ($n = 11$), others did not present compliance data on continuation of treatment ($n = 10$), did not include more than 50 patients ($n = 3$) or ($n = 5$) were concerned with a different patient group (e.g. snorers). One was not available in full form.

Brette *et al.*[192] assessed long-term MAD use in a French cohort ($n = 140$) with mean AHI of 27 events/hour at baseline. The device assessed, 'uses thermoformed splints custom-fitted to the patient's dental arches based on moulds [sic]'. Compliance was determined by a one-off questionnaire at a mean of 2.75 years from treatment initiation, when 76% of patients were still using the device regularly. Vezina *et al.*[193] conducted a retrospective study ($n = 81$) of the use of two different MADs, a traction- and compression-based device, with mean follow-up of 3.6 years. Both devices were custom made from hard copolyester (outer layer) and soft polyurethane (inner layer), following dental impressions. They found that 59% of patients were still using the MADs. Ghazal *et al.*[194] conducted a long-term (mean follow-up of 3.5 years) randomised study of two MADs ($n = 103$). At follow-up, 62% and 46% of patients were still using the two different devices, the first being an IST (hard methylmethacrylate) and the latter a Thornton Anterior Positioner® (made of a laminated, hard–soft polymer with an inner soft polyurethane and an external hard polycarbonate component). In a prospective study with mean follow-up of 1.4 years, which included telephone survey follow-up, Gindre *et al.*[195] reported that 82% of patients ($n = 66$) were still using the device, on average 6 days a week. The majority of this group ($n = 50$) had moderate to severe sleep apnoea (mean AHI = 38.6), but had not been able to tolerate CPAP.

In a real-life study of CPAP compliance ($n = 303$), Galetke *et al.*[196] observed, after a median follow-up of 13 months, that 67% of participants were still regularly using the CPAP machine, while 27% had definitively discontinued use. Mean AHI in this group was 33 events/hour and mean ESS score was 9. A prospective study ($n = 158$) investigating titration methods for CPAP also collected some data on long-term compliance (median follow-up 1.9 years) and found that 77% were still using CPAP at 3 years.[197] Kohler *et al.*[198] conducted a long-term study of usage of CPAP in Oxford with median follow-up of 3.9 years. After 5 and 10 years, 81% and 70% of patients were still using CPAP. They also investigated covariates associated with adherence and found that only ODI was a significant factor, suggesting that more severe apnoea is associated with greater compliance, as McArdle *et al.*[8] demonstrated. However, subjective daytime sleepiness was not a significant factor.

Hoffstein[199] pooled data from 21 studies of MAD compliance, to produce an estimate of 56–68% of patients still wearing the device at 33 months, though some of these patients had very limited symptoms.

Estimates of CPAP compliance from Kohler *et al.*,[198] who conducted a large hospital record-based study of 600 patients in England, were used in our updated modelling. This gave 10-year data compared with the 4-year data from McAardle *et al.*[45] Based on mean AHI of 30 events/hour in the McAardle *et al.*[45] population and mean ODI of 28 events/hour, these groups can be considered to be of broadly similar severity, although mean ESS score is higher in the Kohler *et al.*[198] population. Though some compliance data regarding MADs were identified, the picture is unclear. The assumption that compliance for MADs was the same as for CPAP therefore remained unchanged. There is evidence to suggest that CPAP compliance is lower in milder severity groups, but there is no corresponding evidence that MAD compliance would necessarily be higher. Scenario analyses were therefore conducted to investigate the effect of different compliance rates for CPAP and MADs. Kohler *et al.*[198] estimated a HR of 0.97 for ODI. This means that a fall in ODI of 10 events/hour would represent an increase in risk of discontinuing CPAP therapy of 26%. There are no similar data on the relationship in MADs. Therefore, a one-way conservative adjustment to CPAP compliance was made, reducing it by 5% and 10% to observe the effect.

Mortality rates

Non-cardiovascular disease mortality, originally based on data from 2004 in McDaid *et al.*,[8] was updated using interim life tables (2009–11) and mortality statistics for 2010 from the Office for National Statistics.[200,201] The interim life tables gave age- and gender-specific mortality rates, from which the all-cause hazard was reduced according to the proportion of people who died of CHD and ischaemic heart disease. Underlying mortality rates for patients who have suffered a stroke or CVE were adjusted based on data from two long-term follow-up studies, and are shown in *Table 41*.

Modelling treatment effects

Treatment effects were taken from the meta-analysis presented in *Chapter 3* for mild to moderate OSAH. This analysis suggests that the difference in ESS score for CPAP and MADs are very similar: −1.62 and −1.61, respectively. In a scenario analysis, device-specific differences in ESS score observed in the TOMADO study to estimate cost-effectiveness for the SP1, SP2 and bMADs were used. Differences in BP were also taken from the meta-analysis, though, given the data, it was not possible to estimate specifically for a mild to moderate group. The risk of RTA was based on the CPAP treatment effect pooled by McDaid *et al.*[8] and the ratio of ESS score for MADs and CPAP. These effects are presented in *Table 46*. The base-case risk of RTA after use of MAD is shown, but in scenario analyses will differ according to the ESS treatment effect.

Resource use and costs

McDaid *et al.*[8] incorporated into the model the costs (at 2004/5 prices) relevant to the NHS and personal social services which included the cost of the three interventions (CM, CPAP and MADs) and on-going costs associated with their provision, as well as those of OSAH-related events (RTAs and CVEs).

TABLE 46 Modelled treatment effects

Parameter		Mean difference	SD	Source
ESS	MAD vs. CM (mild to moderate)[a]	−1.620	0.380	Meta-analysis (see *Chapter 3*)
	CPAP vs. CM (mild to moderate)	−1.610	0.340	Meta-analysis (see *Chapter 3*)
SBP	MAD vs. CM	−1.130	0.530	Meta-analysis (see *Chapter 3*)
	CPAP vs. CM	−2.360	0.660	Meta-analysis (see *Chapter 3*)
Risk of RTA	MAD vs. CM	0.167	–	McDaid *et al.*[8] and ratio of ESS treatment effects
	CPAP vs. CM	0.168	0.033	McDaid *et al.*[8]

a Mild to moderate based on mean baseline AHI of study participants.

The cost of CHD events was taken from an evaluation of cardiac medication. Briggs *et al.*[148] used data from a large trial ($n = 12,218$) extrapolated using Markov modelling to estimate 'background' costs as well as the costs associated with modelled events. From regression analyses on costs, McDaid *et al.*[8] were able to utilise the estimated cost for fatal CVEs (which tends to be somewhat lower than for non-fatal events) as well as the cost of an acute CHD event and on-going treatment of CHD. These data were assigned to the health states in the model and to the models for risk of CVEs. Similarly McDaid *et al.*[8] identified a study which would give the acute cost of a stroke and the on-going costs associated with being in a post-stroke health state. Bravo Vergel *et al.*[149] used long-term data from the Nottingham Heart Attack Registry (5 years) which gave details of frequency, timing of recurrent events and in-depth resource use. The costs of RTAs were taken from Department of Transport estimates of the NHS costs associated with fatal and non-fatal RTAs.

For the purpose of this cost-effectiveness analysis, costs were updated where possible and presented at 2011/12 prices. Where relevant, costs were increased for health-care service inflation using PSSRU price indices.[147] The costs of CPAP, MAD and CM are shown in *Table 47*. The cost of a CPAP machine was provided by Meditas and the cost of an auto-adjusting positive airway pressure (APAP) machine used in the titration process by Respironics. Information provided by ResMed in its submission to NICE[37] was taken to estimate the cost of starting CPAP therapy and on-going yearly costs. A survey of clinicians was used to estimate the cost of the titration process based on the proportions that undergo outpatient and inpatient titration and the method used. These data were assessed for face validity by the TOMADO clinical team. Outpatient visits in sleep clinics were updated for contemporary reference costs, as was the cost of specialist nurse time. The acute cost of CPAP therapy was estimated to be £173. Along with other annual costs and the assumption that the lifespan of a machine was 7 years, equivalent annual cost was estimated to be £252.

In the base case, the costs of MADs were assumed to be those of the SP2, as presented in *Chapter 2*. Based on clinical opinion it was assumed that, on average, a patient would have one annual follow-up with a sleep specialist. The lifespan of the device was assumed to be 1 year, based on the expectations of the manufacturer and clinical opinion, as no long-term evidence of replacement was available. Given the comparatively short lifespan and inability to return MADs for reuse, this was noted as a potential source of uncertainty and investigated in scenario analyses, along with using the costs for the SP1 (1-year lifespan) and the bMAD (a fully bespoke MADs assumed to have a lifespan of 18 months).

The costs of CM were taken to include a one-off consultation with a GP. This was taken from PSSRU estimates.[147]

TABLE 47 Costs associated with interventions (2011/12 prices; £)

Cost parameters	Mean	SD	Source
CM	36.00		PSSRU[147]
CPAP initial costs			
Unit cost of follow-up outpatient visit	105.89	47.08	NHS reference costs 2011/12[58]
Probability of having a follow-up outpatient visit	0.69	0.3	McDaid *et al.*[8]
Total cost of follow-up outpatient visit	*73.06*		
Probability of using APAP	0.81	0.19	McDaid *et al.*[8]
Probability of home titration	0.99	0.01	McDaid *et al.*[8]
APAP machine	499.00		Jenny Salmon, Phillips Respironics, 2013, personal communication
Number times CPAP/APAP used for dose titration	163		McDaid *et al.*[8]
Total cost APAP for dose titration	*3.06*		
Probability of using CPAP	0.19		McDaid *et al.*[8]
CPAP machine	230		Angela Dunnil, ResMed UK Ltd, 2013, personal communication
Total cost CPAP for dose titration	*1.41*		
Total cost of in-home titration	*2.72*		
Probability of inpatient titration	0.01		McDaid *et al.*[8]
Unit cost sleep study follow-up	722.80	263.56	NHS reference costs 2011/12[58]
Total cost of inpatient titration	*7.23*		
Probability of seeing a specialist nurse for titration	1		McDaid *et al.*[8]
Unit cost of 30-minute appointment with specialist nurse	44.50		PSSRU[147]
Total cost of specialist nurse involved in titration	*44.50*		
Probability of seeing a consultant for titration	0.4	0.4	McDaid *et al.*[8]
Unit cost of consultant appointment	105.89	47.08	NHS reference costs 2011/12[58]
Total cost of titration by consultant	*42.37*		
Unit cost of 30-minute appointment with technician	11.23		McDaid *et al.*[8] inflated
CPAP initial cost	*174.94*		*(73.06 + 2.72[a] + 7.23 + 44.5 + 42.37 + 11.23)*
CPAP on-going costs			
Interest rate	3.5%		NICE[37]
Estimate life of CPAP machine (years)	7		McDaid *et al.*[8]
Annual equivalent cost CPAP machine	*36.34*		*230/annuity factor*
Cost of CPAP mask	105.00		ResMed (50% full/50% nasal masks)
Estimated life of CPAP mask	1		McDaid *et al.*[8]
Annual equivalent cost CPAP mask	*92.43*		
Annual sundries	*17.33*		*McDaid* et al.[8] *inflated*
Annual follow-up	*105.89*		*NHS reference costs 2011/12*[58]
CPAP on-going annual cost	*251.99*		*(36.34 + 92.43 + 17.33 + 105.89)*

TABLE 47 Costs associated with interventions (2011/12 prices; £) (*continued*)

Cost parameters	Mean	SD	Source
MAD initial costs			
Thermoplastic device (SP1)	*21.00*		*TOMADO*[77]*, Chapter 2*
Semi-bespoke device (SP2)	*128.00*		*TOMADO*[77]*, Chapter 2*
Bespoke device (bMAD)	*552.00*		*TOMADO*[77]*, Chapter 2*
MAD on-going annual cost	*105.89*	*47.08*	*NHS reference costs 2011/12*[58]

a Weighted cost of CPAP/APAP titration.

The costs of CHD and stroke as modelled by McDaid *et al.*[8] were taken from robust long-term data sources. No new sources identified were able to reflect the acute costs of events and on-going costs associated with these conditions in a way that suited the modelling and so these costs were increased for general health service inflation. No new UK-specific estimates of the costs associated with RTAs were identified and so those used by McDaid *et al.*[8] were inflated to reflect 2011/12 prices. These are shown in *Table 48*.

Methods of analysis

The base case includes a hypothetical cohort of 10,000 men informed by the characteristics of the TOMADO population. These characteristics are shown in *Table 40*. ESS treatment effects were taken from the meta-analysis stratified to include studies of OSAH that fell into the mild to moderate range according to mean baseline AHI. Costs were based on the SP2 device, with an assumed lifespan of 12 months. All models incorporated the uncertainty around model input parameters by repeatedly sampling ($n = 15{,}000$) from the parameter distributions and recalculating model outputs conditional on each sample, in order to estimate the distribution of the outputs. Distributions were chosen dependent on the nature of the parameter being sampled. Gamma distributions were used for unit costs, Normal distributions were used for input parameters that were estimated from regression coefficients (including the Cholesky decomposition of mapped utility values) and log-Normal distributions were used for relative risks. Several scenario analyses were conducted, which still incorporated the probabilistic elements of the modelling and, where relevant, adjusted distributions of input parameters accordingly.

TABLE 48 Mean costs associated with CHD, stroke and RTAs

Cost	Mean	SD	Source
CHD and stroke			
Cost of fatal CVE	3561	434	Briggs *et al.*[148]
Acute cost of CHD	11,786	505	Briggs *et al.*[148]
Ongoing cost of CHD	886	138	Briggs *et al.*[148]
Acute cost of stroke	10,476	347	Bravo Vergel *et al.*[149]
Ongoing cost of stroke	2764	334	Bravo Vergel *et al.*[149]
RTA			
Cost of RTA (non-fatal)	3120	1942	Department of Transport[146]
Cost of RTA (fatal)	6297	1942	Department of Transport[146]

Scenario analyses were conducted to investigate sensitivity of outputs to:

- the lifespan of the interventions
- the cost of devices incorporating SP1 and bMAD costs
- ESS treatment effects observed in TOMADO
- reduced CPAP compliance in lower severity disease using a multiplier
- the time horizon
- use of an alternative source of the relative risk of vascular events given a reduction in SBP
- use of an alternative source for the effect of effective treatment of OSA on RTA events.

All results are presented as incremental cost per QALY. For the base case, uncertainty in the estimates is presented as the likelihood of being cost-effective at WTP thresholds of £10,000, £20,000 and £30,000 per QALY, the CEAC for a range of WTP thresholds and the CEAF to identify the most cost-effective treatment option over the range of WTP thresholds. All costs are in 2011/12 prices.

Results of the economic model

Base-case analysis

The results of the base case are presented in *Table 49*. This shows that MADs compared with CM are more costly but also more effective in patients with mild to moderate OSAH. The additional costs are a result of much higher treatment costs, with a reduction in RTA and CVE costs mitigating this difference somewhat. The ICER of MADs compared with CM is £6639 per additional QALY gained. CPAP compared with MADs is more expensive but more effective. The ICER of CPAP compared with MADs is £14,012 per QALY gained.

At a threshold value of £20,000/QALY, CPAP has the highest mean INMB compared with CM (£3879) and the probability that CPAP is cost-effective is 0.52. At a threshold value of £30,000/QALY, this probability increases to 0.55 with a mean INMB of £6914. Oral devices have a mean INMB compared with CM of £3794 at a threshold value of £20,000/QALY and the probability that they are cost-effective is 0.47. At a threshold value of £30,000/QALY, this probability decreases to 0.45 with a mean INMB of £6643.

TABLE 49 Cost-effectiveness results (base-case analysis)

Cost-effectiveness component	CM	MAD	CPAP
Intervention costs (mean)	£36	£3206	£3524
RTA costs (mean)	£1963	£713	£716
CVE costs (mean)	£4118	£4103	£4074
Total costs	£6116	£8022	£8307
Total QALYs	14.336	14.621	14.640
ICER (oral devices compared with CM and CPAP compared with MADs)		**£6687**	**£15,367**
Probability of cost-effectiveness			
At £10,000/QALY	0.16	0.46	0.38
At £20,000/QALY	< 0.01	0.47	0.52
At £30,000/QALY	0	0.45	0.55

Figure 32 depicts the uncertainty surrounding decisions of which approach is most cost-effective in the base-case analysis, for a range of values decision-makers may be willing to pay per QALY gained. It shows that at very low WTP thresholds, CM is the most likely to be cost-effective. Over the conventional range of £20,000–£30,000, CPAP has the highest likelihood of being the most cost-effective, with the decision becoming less uncertain as WTP per QALY increases. At a WTP of approximately £20,000/QALY the probability that CM is the most cost-effective falls to zero.

Figure 33 gives the CEAF for the base case. It shows the intervention which yields the highest mean net benefit over the range of WTP. It can be seen that, while MADs have the highest mean net benefit after a threshold of £6687, there does remain uncertainty of whether or not it is likely to be more cost-effective than CM. From £15,367, CPAP becomes cost-effective, and at this point the likelihood of MADs and CPAP being cost-effective is very similar, 0.48 and 0.49, respectively. At higher WTP, CPAP always has the highest mean net benefit and highest likelihood of being the most cost-effective, although with considerable uncertainty.

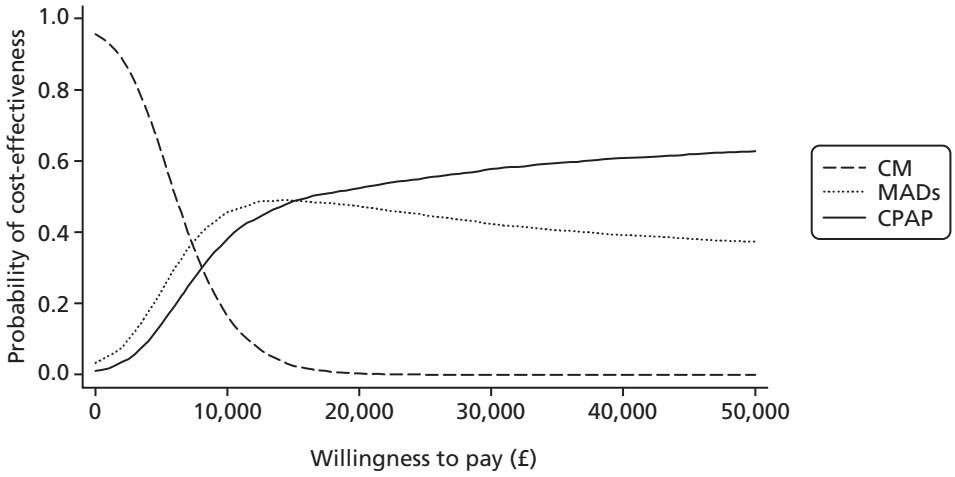


FIGURE 32 The cost-effectiveness acceptability curves (base-case analysis).

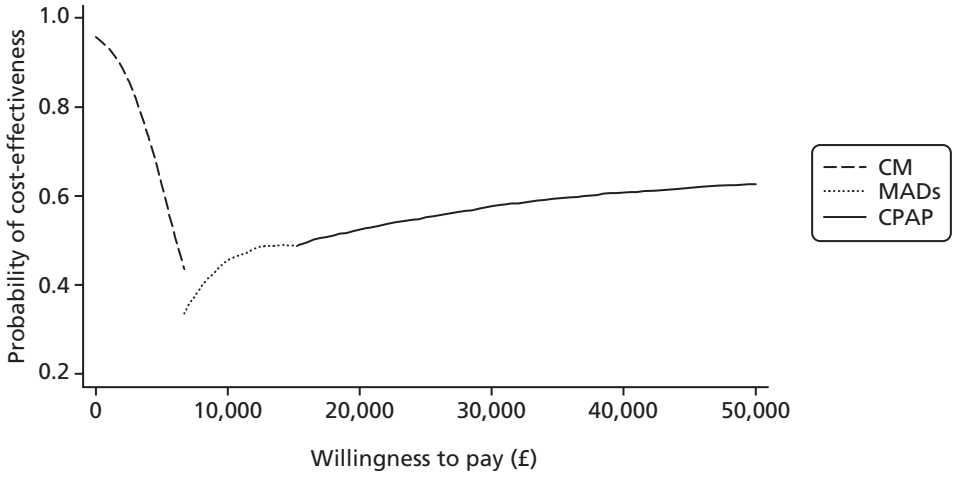


FIGURE 33 The cost-effectiveness acceptability frontier (base-case analysis).

Sensitivity analyses

A series of one-way deterministic sensitivity analyses were undertaken to explore the additional impact of changing specific input values on the cost-effectiveness results from the base case. These are presented in *Table 50*, which shows that decisions are not sensitive to the use of SF-6D utilities scores. This is also true for use of an alternative source of relative risk reduction associated with decreasing SBP. However, results and decisions are sensitive to assumptions about costs. For example, replacing device costs from SP2 with those for SP1 or bMAD costs leads to a different decision about the relative value of CPAP; in the case of SP1, CPAP would no longer be considered cost-effective (ICER = £89,182) by usual NICE threshold values, as the additional benefits of CPAP become relatively more expensive. Replacing SP2 device costs with bMAD leads to CPAP dominating bMAD as the benefits of CPAP are greater and costs are lower than bMAD. This is the case even if the lifespan of bMAD is assumed to be 2 years rather than 18 months.

The assumed lifespan of devices makes a difference to the optimum decision. A conservative estimate for the lifespan of the SP2 based on manufacturer and expert clinical opinion was 1 year. However, if the lifespan is increased to 18 months, SP2 becomes the most cost-effective intervention.

Use of device-specific costs and effects as observed in TOMADO indicates that SP2 dominates CPAP, given the comparatively higher QALYs gained. A comparison of bMAD with CPAP shows that both the costs and benefits of CPAP are lower. However, at a conventional threshold of £20,000 to £30,000 per QALY,

TABLE 50 Summary of ICERs following deterministic sensitivity analyses

	Type of deterministic sensitivity analysis	
	MADs vs. CM	CPAP vs. MADs
Base case	***£6687***	***£15,367***
Length life SP2 12 months – > 18 months	£4674	£44,066
Utility derivation		
EQ-5D-3L – > SF-6D QALYs	£8783	£16,225
Relative risk reduction for CVE associated with unit fall in SBP		
Reduction in cardiovascular risk associated from Lewington *et al.*[174]	£6741	£14,606
MAD costs		
SP1 device costs (assuming 12-month lifespan)	£1552	£89,182
bMAD costs (assuming 18-month lifespan)	£18,161	Dominant
bMAD costs (assuming 2-year lifespan)	£13,836	Dominant
TOMADO device-specific costs and treatment effects		
SP1 costs (12-month lifespan) and effects (ESS = –1.51)	£1656	£56,640
SP2 costs (12-month lifespan) and effects (ESS = –2.15)	£5425	Dominated
bMAD costs (18-month lifespan) and effects (ESS = –2.37)	£14,539	£57,907
Time horizon		
10-year time horizon	£8309	£90,998
RTA treatment effect		
Treatment effect from Tregear *et al.*, 2010[180] meta-analysis	£17,002	£16,428
Compliance		
CPAP compliance reduced by 5%	£6667	£40,668
CPAP compliance reduced by 10%	£6756	Dominated

the cost savings of CPAP compared with bMAD are larger than the value to 'compensate' for lower benefits of CPAP.

If a shorter time horizon is considered, CPAP becomes less cost-effective. This is because much of the benefit of CPAP results from its greater effectiveness in lowering BP. The benefits of reducing this risk factor for CVD would accrue later in patients' lives.

Summary and discussion

This chapter builds on a well-developed existing economic model, to assess the cost-effectiveness of MADs compared with CPAP and CM for patients with mild to moderate OSAH. Updated and new reviews of the evidence were conducted to reflect evidence that has emerged since the original modelling exercise and to better represent patients with mild to moderate OSAH. These covered the role of sleep apnoea in CVD, RTAs, HRQoL and long-term compliance by treatment.

Understanding of the mechanism of sleep apnoea on CVD has developed since the original model and, despite some conflicting evidence, the body of published studies indicates probable causality. However, there are still no reliable long-term data on cardiovascular outcomes under different treatment options for sleep apnoea. The model relies on differences in BP as proxies, reflected through prediction of risk using the Framingham equation, and direct evidence would improve the modelling. Data on BP from trials are heterogeneous and there are insufficient data to separate the effects by severity of disease. Data from new meta-analyses on the risk of RTA were used in sensitivity analysis rather than the base-case analysis because of difficulties in ascertaining the reasons for inclusion of papers. The use of generic measures of HRQoL in randomised trials to support conversion onto a utility scale is still rare, but TOMADO enabled a re-estimation of the relationship between ESS score and utility based on more data and for different levels of severity. The literature search for compliance data identified the longest-term follow-up study of CPAP compliance to date, but similarly robust data are still not available for MADs.

The meta-analysis presented in *Chapter 3* fed into the model and, by estimating a similar treatment effect for MADs and CPAP, indicates the likely importance of the cost of delivering the treatment options. The base-case analysis for MADs used trial data from *Chapter 2* based on the cost of SP2, with sensitivity analyses focusing on the cost of SP1 and bMAD as well as the length of life of the device. The costs of CPAP and CM were based on inflation-adjusted estimates from McDaid *et al.*[8] supplemented by company-supplied prices.

The results from the updated model suggest that, at conventional NICE thresholds of £20,000 to £30,000 per QALY, both MADs and CPAP are cost-effective compared with CM. CPAP is the preferred option, at a WTP per QALY of £15,000 and above. However, there is considerable uncertainty with CPAP having a 52% probability of being the most cost-effective option at £20,000 per QALY, compared with 47% for MAD. As cost per QALY increases to £30,000, the corresponding figures are 55% for CPAP and 45% for MAD. These suggest that MADs could be considered a legitimate treatment option for mild/moderate sleep apnoea, especially if CPAP is not tolerated.

The sensitivity analyses indicate that the cost of devices and their lifespan is important for the policy decision. For example, assuming costs for the bMAD, rather than the SP2, results in the CPAP being both more effective and less costly even with a 2-year lifespan for the bMAD. However, increasing the length of life of the MAD from 12 months to 18 months, or using SP1 costs in place of those for SP2, results in an increase in the incremental cost per QALY for CPAP relative to MAD to £44,066 and £90,998, respectively. Long-term data on the lifetime of MADs in routine practice would improve precision of estimates.

The sensitivity analysis also indicated the importance of compliance. Reducing compliance with CPAP by 5% increases the ICER of CPAP relative to MADs to £40,000/QALY. A reduction of 10% in compliance with CPAP means that QoL gains for CPAP over MADs are lost and the cost is higher. As there is evidence that, for milder sleep apnoea, compliance with CPAP falls and, therefore, that MADs may be more cost-effective, comparable compliance data for MADs are required to confirm or refute this.

Finally, the sensitivity analysis indicates the importance of the time frame of the analysis. Moving from a lifetime to a 10-year time horizon changes conclusions with respect to the relative value of CPAP and MADs; the cost per QALY of CPAP increases from £15,000 to £91,000 per QALY. This is largely because the cost of CPAP is not spread over a sufficiently long period and the value of the increased benefits (e.g. reduced CVD) is not accounted for.

APPENDIX E: Paper 5

Paper 5: HTA impact review

Raftery J, Hanney S, Greenhalgh T, Glover M, Blatch-Jones A. Models and applications for measuring the impact of health research: Update of a systematic review for the health technology assessment programme. Health Technol Assess. 2016;20(76):1-282. https://doi.org/10.3310/hta20760

Extract presented here is: 'Chapter 5: Estimating the monetary value of the impact of health research'

Chapter 5 Estimating the monetary value of the impact of health research

Introduction

The economic impacts from medical research form a subset of many of the logic models presented in *Chapter 3*. A section of the literature has addressed the specific issues relating to undertaking exercises to determine economic impacts or the returns on investment from medical research and development spending. Previous reviews of the literature, which form a starting point for this review, have highlighted work that has been done to advance the field.[2,205] The methods used to assess these impacts or returns on investment are born from the economic evaluation literature, and the difference in approaches lies largely in the scope of the cost and benefits assessed, and the valuation methods for seemingly non-monetary components of the impact. As stated in *Chapter 2*, the purpose of this review was, using Buxton *et al.*[28] as a starting point, to identify studies since 2004 that have used any methods to attempt to value (in monetary terms) the benefits (health and cost savings) of a body of health research and link that with an investment in the body of research. Articles were included only if they contained a component that attempted to value the impact of research and development investment on population health.

The article in the *Bulletin of the World Health Organization* by Buxton *et al.*[28] attempted to learn from previous studies that had estimated 'monetary values for the societal benefits obtained from health research, especially those studies that have attempted to link (and value) benefits to a specific society from a specified (and costed) body of research'.

The authors characterised the identified methods into four categories:

1. valuing direct cost savings to the health-care system
2. valuing benefits to the economy from a healthy workforce
3. valuing benefits to the economy from commercial development
4. measuring the intrinsic value to society from health gain.

Studies were identified that had considered the benefit of medical research and development as direct cost savings to the health-care system, brought about by a reduced number of people requiring treatment or reductions in per patient treatment costs. This approach had been predominant in estimating the benefits of vaccination research, which had the potential to eradicate subsequent disease and associated treatment costs.[205–208] Cost savings could be included as part of cost–benefit analysis, but these studies did not always link this to an investment period or country-specific research.

One of the earliest studies to attempt to calculate a rate of return from medical research was conducted by Mushkin and Landefeld.[209] A human capital approach (equating the value of life to market values, i.e. wages) was used to value gains from US biomedical research, characterised by a healthier workforce. The limitations of such an approach were acknowledged by the authors and others[28,209,210] and tend to overstate benefits when lost labour can be replaced, while understating benefits for those sections of the population not of working age.

Buxton *et al.*[28] drew largely on a review conducted by Salter and Martin,[211] which explored the commercial economic benefits from basic research. Salter and Martin noted progress made by Mansfield[212,213] that estimated a worldwide social rate of return (benefits accrued to the whole of society, as opposed to one firm or funders of one project) of 28% for research undertaken 1975–78. Studies have also demonstrated the economic benefits of medical research through industrial applications to other industries.[214]

An emerging field highlighted by a number of studies in the Buxton *et al.*[28] review had measured the intrinsic value of health gains brought about by research and development. A US initiative of the Mary Woodard Lasker Charitable Trust, Funding First,[215] produced a series of papers that formed a subsequent book.[216] An informal approach used willingness-to-pay methods to value the increased longevity of life experienced by the US population, attributing a fraction of these gains to medical research. The results suggested 'exceptional returns' of nearly 20 times the investment in US medical research. This type of analysis was performed in a more systematic fashion in an Australian study, taking a similar 'top-down' approach to valuing health gains, to produce an estimate of the annual rate of return to investment in research and development.[217] They estimated a favourable benefit-to-cost ratio of 2.40 (i.e. AUS$1 invested creates an additional AUS$1.40 benefit); however, this work has been subsequently criticised because the time for investment in medical research to produce health gains was not considered.[218]

Buxton *et al.*[28] noted that there is significant scope for these methods to be extended and refined to allow more robust estimation of the economic benefits from medical research. In particular, a widely acknowledged central challenge that must be addressed in this kind of analysis relates to the attribution problem; the relationship between investment in research and health outcomes.[15,22,28,115] This manifests itself as several related issues regarding the contribution of health research in improving health outcomes and what would have happened without research, that is the unobservable counterfactual. Assumptions must be made regarding the share of health gains attributable to health research, and given there is an international pool of health research, the contribution of any particular country to particular health gains. Finally, assumptions must be made regarding the temporal relationship between a period of investment and a period of health gains. Different approaches face somewhat different problems in dealing with attribution, but methods have continued to be developed to address these issues.

Review findings

The search of databases produced 413 articles, which were initially screened by a reviewer by title (*Figure 8*). After initial screening and deduplication, 102 articles were screened by abstract. Seventeen articles were reviewed in full, with five included.[22,26,75,218,219] Two of these articles were included in the main literature review.[26,75] One additional report that was not picked up by the supplementary search was included from the main literature review.[25] One additional article and one report known to the authors was also added.[24,220] One in-press article that the authors kindly gave us access to was also included.[27] In total, the review produced nine articles/reports. The studies and methods of assessing return on investment, that included a component that attached a monetary value to health gains, are summarised in *Table 16* (see *Appendix 6*).

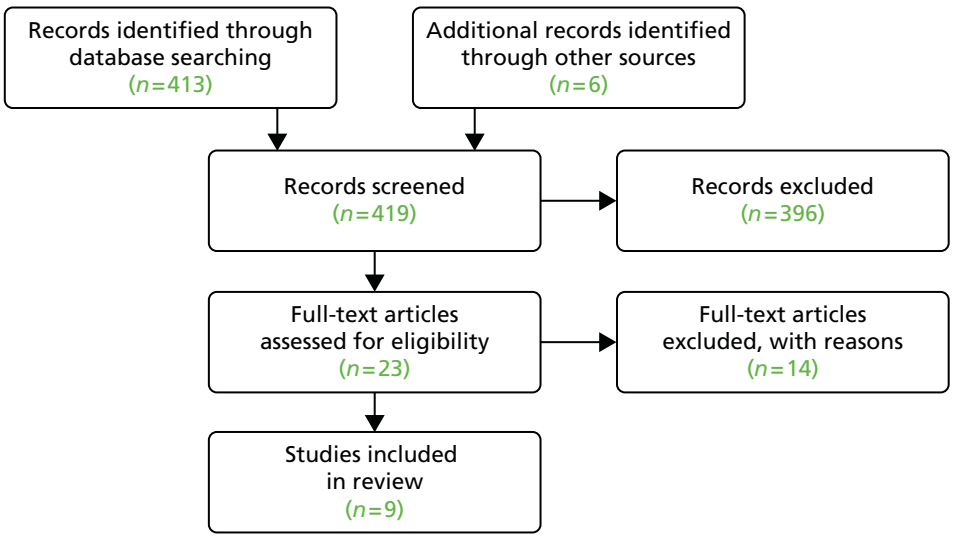


FIGURE 8 Flow diagram of included studies.

The nine studies identified in the review can be split into two categories in terms of how health gains are measured: those taking a top-down approach and those taking a bottom-up approach. There are several other important issues that must be addressed in quantifying the returns; this simple taxonomy allows us to explore the different methodologies. *Figure 9* depicts the basic methodologies.

Studies have been compared on a number of key facets of the analysis and assumptions that have to be made regarding measuring and valuing net health gains and how to attribute a proportion of health gains to a body of research as follows: How were health gains measured? How were health gains valued? Were health gains ex post or ex ante? Were the costs of delivery accounted for? Was the lag between investment and health gain considered? How was the attribution problem addressed?

Top down

A stream of work undertaken by Access Economics (now Deloitte Access Economics) assessed the benefits of medical research in terms of the intrinsic value of the health gains to society. Two studies were conducted to estimate the returns on investment from Australian research and development.[24,25] Access Economics considered all Australian health research and development spending both public and private between 1992 and 2005.[24] Building on their approach in an earlier report,[217] they used projections from the Australian Institute of Health and Welfare to estimate DALYs averted in the period 2033–45 relative to 1993 levels and calculate a return on investment of 117%.[24] The authors assume that the lag between investment and realisation of health gains is 40 years, although the rationale for this figure is unclear.

To calculate the return on investment the authors considered the proportion of DALYs averted attributable to research and development, as opposed to other factors claimed not to be a result of research and development. The authors state that other factors include 'public health awareness and preventive programs such as 'Slip Slop Slap' or 'Quit', screening and early intervention initiatives, the public subsidy of drugs and interventions through the Pharmaceutical Benefits Scheme and the Medicare Benefits Schedule, and so on'.[24] The extent to which these examples are not research and development-based interventions could be heavily debated, especially screening programmes; however, the premise that external factors other than research and development are responsible for health gain has been widely acknowledged.[221] They attributed 50% of health gains to research, as they had in their previous study, but have acknowledged that this was not robust.[217] The return was highly sensitive to the value of this parameter. The authors take account of research and development conducted in others countries and its contribution to Australian health gains by using bibliometric techniques to estimate a proxy, based on Australia's share of publications in the clinical sciences. They estimated that 3% of health gains could be attributed to Australian research and development. The DALYs averted were monetised using a willingness-to-pay methodology, attaching the value of a statistical life-year [AUS$266,843 – £124,300 (converted at 2015 purchasing power parity exchange rate)].[222]

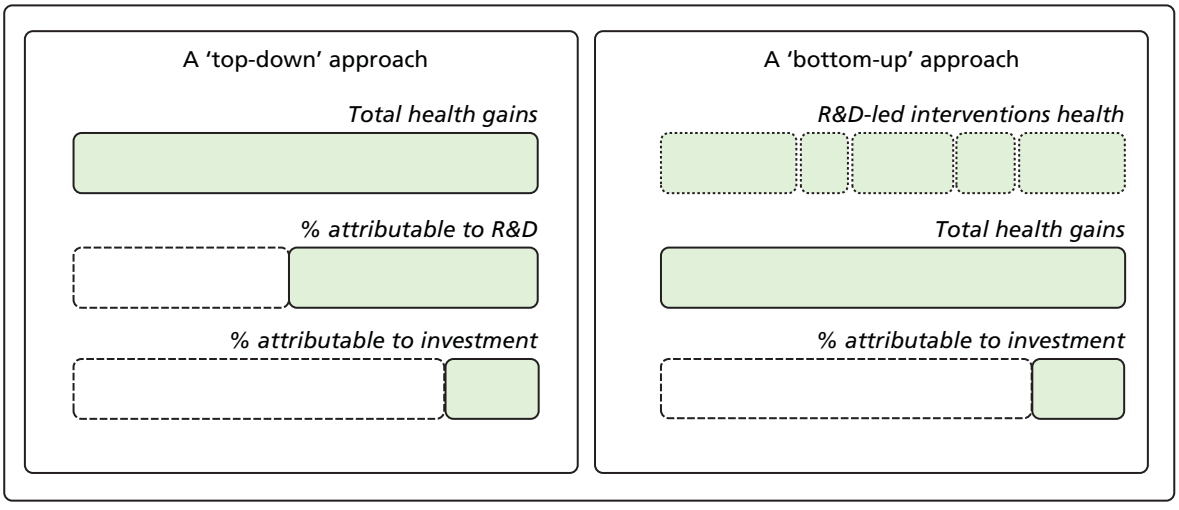


FIGURE 9 Approaches to identifying health gains from research. R&D, research and development.

A further study estimated the returns from National Health and Medical Research Council (NHMRC) funding in five disease areas (cardiovascular disease, cancer, sudden infant death syndrome, asthma and muscular dystrophy) using the same methodology.[25] They estimated that the return on investment ranged from 509% in cardiovascular research to –30% for muscular dystrophy. These returns also included the value of avoiding direct health system expenditure, the value of avoiding indirect costs (through productivity losses), the value of direct commercial gains from the NHMRC-funded research and development, and benefits of NHMRC-supported commercialisation. Neither of the Deloitte Access Economics studies considered potential increased costs borne by the health-care system from expensive new technologies.[24,25]

Health gains were measured using DALYs averted in 2040–50 relative to 2000 levels.[25] The time between investment (2000–10) and health gains (2040–50) was again assumed to be 40 years. It was assumed that the proportion of gains as a result of research and development was 50%, and 3.14% of these gains were assumed to be attributable to Australian research and development (re-estimated using bibliometric techniques). However, the authors were presented with an additional necessary estimation; the proportion of health gains that were a result of this programme of NHMRC research, rather than the whole body of Australian health research and development. Using bibliometric techniques they found that 25.04% of Australian research publications were funded through the NHMRC and used this as a proxy. The DALYs averted were monetised using a willingness-to-pay methodology (based on individual's valuation of avoiding mortality/morbidity) attaching a value of a statistical life-year [AUS$168,166 – £78,300 (converted at 2015 purchasing power parity exchange rate)].[222]

Roback *et al.*[218] used a broadly similar approach to value Swedish gains from all public and private research and development spending on clinical and health research in the year 2005. In this tentative modelling exercise, average annual increases in life expectancy (population utility adjusted) were used to estimate QALY gains in 2015. This implies a lag of 10 years, but the authors did not explicitly discuss this. QALY gains were valued using the value of a statistical life-year [SEK500,000 – £37,900 (converted at 2015 purchasing power parity exchange rate)].[222] The returns were estimated at a socioeconomic level, including a whole range of non-health benefits where they could be quantified, resulting in a return on investment of 1.08 (8%). In making this estimate, the authors did account for 'more expensive healthcare due to new methods'. They assumed that 50% of health gains were attributable to research and development, referencing estimates made by various authors that suggest the range may be between 25% and 67%.[216,223–225] The proportion of health gains attributable to Swedish research and development was assumed to be 3% based loosely on an estimate of Sweden's share of global expenditure and global medical publications.

A significant drawback of Roback *et al.*[218] and the Access Economics[24,25,217] work is the ex ante nature of health gains: the reliance on predictions based on previous trends in population health improvement. This assumes the impact of as of yet unobserved future usage of interventions and hence improvements in health. Ex post studies use retrospective data, either by directly observing population health gains or by compiling data on observed uptake and modelled per patient incremental net health benefits. Although many of these studies require pragmatism in assumption making, the reliance on unknown unknowns requires a leap of faith.[24,25,217,218]

Lakdawalla *et al.*[220] assessed the social surplus arising from the 'war on cancer' in the USA from all public cancer research and development spending between 1971 and 2000. An upper bound of this investment was estimated to be US$300B, based on National Cancer Institute spending (which was assumed to make up approximately one-quarter of cancer research and development spending). Ex post life-year gains in survival between 1988 and 2000 were identified and valued at individual willingness-to-pay [US$30,737 – £21,300 (converted at 2015 purchasing power parity exchange rate)].[222] This produced an estimate of the net gains at US$1.6T. Lakdawalla *et al.* acknowledge the likely lag between investment and health gain and suggest that they may have overestimated the size of investment and hence conservatively estimate social surplus, but did not explicitly investigate the lag. The survival gains were estimated based on cancer-specific improvements in detection and treatment, although the potential for non-research and development contributions to these improvements was not considered.

Bottom up

Informed by methodological frameworks, such as the Payback Framework, studies have used a different approach to build the benefits up from individual interventions to estimate the sum of the health gains, rather than starting from an estimate of overall health gains.[39] It, in part, theoretically deals with the attribution problem presented when trying to estimate the contributions of research and development and non-research and development factors in producing health gains, although it produces a different challenge in identifying only those interventions that are known to have been research driven.

Johnston *et al.*[75] applied such an approach to the US National Institute of Neurological Disorders and Stroke's funding of 28 Phase III RCTs prior to 2000. They estimated a return on investment of 46% per year based on 10-year estimates of post-funding QALYs. Available cost–utility analyses were used to estimate the per-patient QALY gains for eight interventions, and data on use were gathered to estimate population gains. Implicitly, it was assumed that all changes in use post trial were a result of that clinical research. Although the examination of the use of the eight interventions suggests some lag, with use fairly stable for at least 2 years after the completion of funding, it might be considered shorter than other estimates.[226] Data presented by the authors suggest that use is not zero during the period when funding ends, which might be indicative that other research not funded by the National Institute of Neurological Disorders and Stroke could have played a role in health gains. By using cost per QALY utility data, the authors were able to present monetised health gains [valued at GDP per head of US$40,310 – £27,900 (converted at 2015 purchasing power parity exchange rate)][222] net of costs of delivery (net monetary benefits) for each intervention. The study was able to find adequate data for only 8 of the 28 Phase III trials, which highlights the data-heavy nature of this exercise and the reliance on published literature. In some instances, a paucity of data may limit the ability for such a study to be undertaken or at least limit the generalisability of findings.

Two studies published by authors of the Payback Framework have adopted an approach that is similar with respect to the identification of health gains to the work of Johnston *et al.*,[75] but have focused on quantifying the returns in different disease areas.[22,26]

Buxton *et al.*[22] estimated the return on investment [presented as an internal rate of return (IRR) that considers the flow of cost and benefits] from publicly and charitably funded cardiovascular research in the UK to be 9% per year (£1 investment yields health gains equivalent to £1.09). They estimated the health gains between 1986 and 2005 and linked this with a period of investment between 1975 and 1988, based on a lag of 17 years. The lag was estimated based on citation analysis of UK guidelines, using mean time between citation and guideline publication ('knowledge cycle time') as a proxy for the time between investment and health gain. Research-led interventions in the cardiovascular field were identified and a timeline of usage assembled. For each of the interventions, per-patient QALY gains and net costs (increases from delivery and potential savings from reduced sequelae) were identified through published cost–utility analyses. QALY gains were valued at the health-care service opportunity cost based on implied cost-effectiveness thresholds of NICE (£25,000) and presented net of costs, to produce an estimate of the net monetary benefits produced per year. The NICE threshold value was chosen to reflect the competing nature of funding of health research over provision of existing technologies. It was assumed that 17% of the health gain was attributable to UK research, based on bibliometric analysis of cardiovascular guidelines that identified the proportion of cited work that contained a UK corresponding author. Buxton *et al.*[22] combined this IRR with the wider GDP spill over effects of research and development, estimated to be 30%, to give an overall IRR of 39%.

Glover *et al.*[26] applied the same methodology to publicly and charitably funded cancer research in the UK, re-estimating the lag between investment and health gain and the proportion of health gains attributable to UK research based on cancer guidelines. An IRR of 10% was estimated, based on the monetised net health gains for 1991–2010 for research-driven interventions, linked to cancer funding between 1976 and 1995 (15-year lag). This work highlighted the difficulty in identifying all of the important research-driven interventions. An additional publication[227] used accompanying case studies to highlight the complex and

heterogeneous relationship between research and health gains. There is a need in a field such as cancer to narrow the scope to complete such a resource-intensive exercise, where there have been widespread improvements in detection and treatment brought about by research, and where the benefits are realised across a heterogeneous patient population (for instance there are over 200 types of cancer). Although developing a method that used changes in incidence and survival gains as a predictor of which cancer types were likely to have contributed largely to overall gains, the authors assumed that interventions not represented in the analysis produced zero net benefit.

A study by de Oliveira *et al.*[219] largely replicated the methods presented in Buxton *et al.*[22] to assess the return from Canadian publicly and charitably funded cardiovascular research, which they estimated to be 21% per year based on QALY gains in 1994–2005. Using similar bibliometric techniques, a time lag of 13 years was estimated and 6% of overall health gains were attributed to Canadian research and development. They also argued that an additional component should be considered as part of the attribution problem, assuming that 70% of the health gains were attributable to medical research. However, if the identified interventions were research led and studies used to estimate per patient health gains produced incremental differences brought about by the specific intervention, it is not clear why non-research and development factors ought to be considered in this context.

Guthrie *et al.*[27] estimated the benefits of the NIHR HTA programme funding from 1993 to 2013. They selected 10 key HTA studies, which were largely made up of randomised trials but also systematic reviews. They identified the per-patient QALY gains associated with the interventions. QALY gains were monetised at the health-care opportunity cost (£20,000 and £30,000) net of health service costs, but total actualised gains were not estimated. Instead, a net monetary benefit associated with a hypothetical 1 year of full implementation for the patient population of the interventions was calculated; therefore, the lag between investment and gains was not considered. The HTA studies were considered to be responsible for all post-HTA research implementation, as they were seen to constitute 'definitive' evidence. The authors suggest that only 12% of potential net benefit would cover the £367M invested by the NIHR HTA programme. Although indicative of potential gains, this analysis does not adequately address the attribution problem and makes no consideration of when benefits accrue. It also raises the interesting problems posed when the research takes the form of systematic reviews and the role of such a study in changing clinical practice and hence leading to health gain.

Discussion

There have been contributions to the literature that estimate the impacts of health research using methods to attach a monetary value to health gains. Approaches have attempted to estimate the resultant health gains from investment in bodies of research, and, in doing so, must deal with several problems relating to attributing health gains to particular investments. Techniques that attempt to deal with the problems of attribution have been established. However, authors have acknowledged a simplification of the relationship that is required and the reliance on a logic model view of research impacts. Some of these contributions also consider non-health sector benefits falling on the wider economy, although the scope of the benefits considered often differs, as does the valuation.

Only a few studies specifically considered programmes of health research.[25,27,75] Guthrie *et al.*[27] estimated the gains of the NIHR HTA programme, but made cautious conclusions on the returns based on hypothetical uptake of a subset of HTA-funded research.[27] Clearly there is scope for these types of methods to be applied to estimate returns from programmes such as NIHR HTA, but several additional considerations need to be taken into account. Conversely, there are advantages to having a well-defined unit of analysis.

It would appear that assessing monetised impact at a programme level is conducive to the bottom-up approach, when the set of interventions is well defined and the task of identifying those that are

'important' could be avoided. To an extent, data feasibility issues that limit the bottom-up approach should be mitigated by programmes such as NIHR HTA in that most of its research includes cost–utility estimates. However, issues of scale are present if the number of studies undertaken by a programme is large, such as in the NIHR HTA programme. This might be mitigated to an extent by the need to consider only those trials that showed a significant effect, but this makes a bold assumption about the nature of evidence being used in clinical practice.

When attempting to measure health gains from a programme of research, the attribution problem manifests itself as an added layer of uncertainty regarding the proportion of total health gains that should be attributed to the specific programme. Using a top-down approach, Deloitte Access Economics[25] dealt with this by using the percentage of total citations in clinical sciences that were studies funded by the programme as a proxy. This additional attribution problem is not circumvented by the bottom-up approach and a consideration must be still be made. The view taken in Johnston *et al.*[75] and Guthrie *et al.*,[27] that National Institute of Neurological Disorders and Stroke's trials and HTA studies are definitive in terms of changes in uptake, is insufficient, especially in developed countries with multiple funding streams and complex and evolved research ecosystems. The use of the weights attached to particular RCTs in meta-analyses could provide a more systematic way of considering the relative impact of different clinical research supported by multiple funders. Regarding the proportion of health gains that should be attributed to world research and development, this could be used as an intervention-specific replacement for guideline analysis or used as an adjunct. Although time lags must be included, it is not clear how best to estimate these.

An additional problem for programmes that fund only clinical research is dealing with the role of basic research in health gains. This is unclear and constitutes a major potential limitation in these methods. No study looking at a programme has yet encompassed these kinds of considerations into the approach dealing with attribution of health gains to a programme, using either top-down or bottom-up methods.

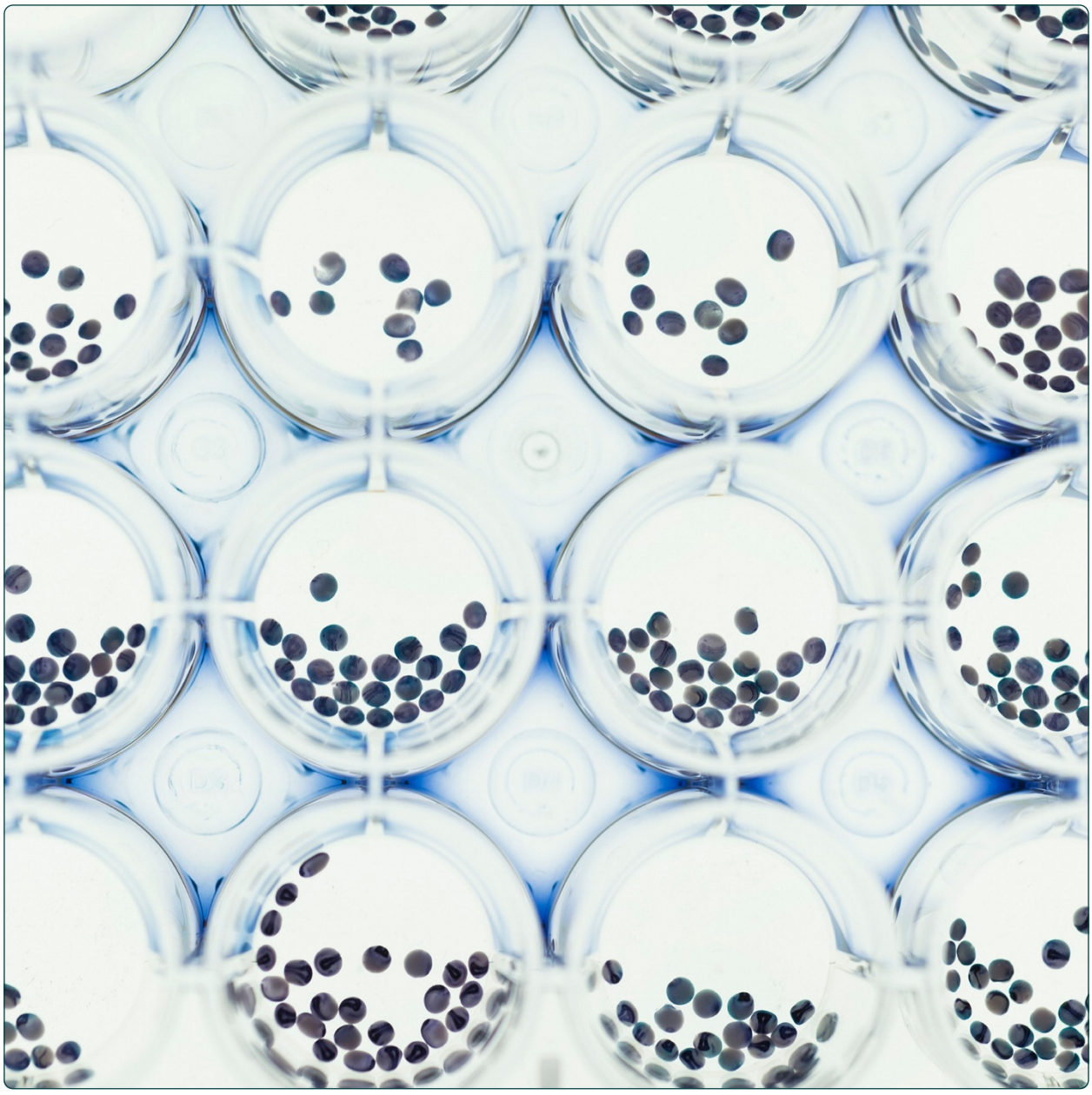
Although it would clearly be possible to estimate the returns on investment from the NIHR HTA programme, significant challenges remain.

APPENDIX F: Paper 6

Paper 6: Estimating returns of cancer research

Glover M, Buxton M, Guthrie S, Hanney S, Pollitt A, Grant J. Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. BMC Med 2014 121. 2014;12(1):1-21. https://doi.org/10.1186/1741-7015-12-99

BMC Medicine



Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes

Glover *et al.*

BioMed Central

Glover *et al. BMC Medicine* 2014, 12:99
http://www.biomedcentral.com/1741-7015/12/99

Glover et al. BMC Medicine 2014, 12:99
http://www.biomedcentral.com/1741-7015/12/99

BMC Medicine

RESEARCH ARTICLE Open Access

Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes

Matthew Glover[1], Martin Buxton[1], Susan Guthrie[2], Stephen Hanney[1], Alexandra Pollitt[2] and Jonathan Grant[2,3*]

Abstract

Background: Building on an approach developed to assess the economic returns to cardiovascular research, we estimated the economic returns from UK public and charitable funded cancer-related research that arise from the net value of the improved health outcomes.

Methods: To assess these economic returns from cancer-related research in the UK we estimated: 1) public and charitable expenditure on cancer-related research in the UK from 1970 to 2009; 2) net monetary benefit (NMB), that is, the health benefit measured in quality adjusted life years (QALYs) valued in monetary terms (using a base-case value of a QALY of GB£25,000) minus the cost of delivering that benefit, for a prioritised list of interventions from 1991 to 2010; 3) the proportion of NMB attributable to UK research; 4) the elapsed time between research funding and health gain; and 5) the internal rate of return (IRR) from cancer-related research investments on health benefits. We analysed the uncertainties in the IRR estimate using sensitivity analyses to illustrate the effect of some key parameters.

Results: In 2011/12 prices, total expenditure on cancer-related research from 1970 to 2009 was £15 billion. The NMB of the 5.9 million QALYs gained from the prioritised interventions from 1991 to 2010 was £124 billion. Calculation of the IRR incorporated an estimated elapsed time of 15 years. We related 17% of the annual NMB estimated to be attributable to UK research (for each of the 20 years 1991 to 2010) to 20 years of research investment 15 years earlier (that is, for 1976 to 1995). This produced a best-estimate IRR of 10%, compared with 9% previously estimated for cardiovascular disease research. The sensitivity analysis demonstrated the importance of smoking reduction as a major source of improved cancer-related health outcomes.

Conclusions: We have demonstrated a substantive IRR from net health gain to public and charitable funding of cancer-related research in the UK, and further validated the approach that we originally used in assessing the returns from cardiovascular research. In doing so, we have highlighted a number of weaknesses and key assumptions that need strengthening in further investigations. Nevertheless, these cautious estimates demonstrate that the returns from past cancer research have been substantial, and justify the investments made during the period 1976 to 1995.

Keywords: Medical research investment, QALYs, Cancer, Medical research charities, Value of health, Rate of return, Time lags, Research payback

* Correspondence: jonathan.grant@kcl.ac.uk
[2]RAND Europe, Westbrook Centre, Milton Road, Cambridge CB4 1YG, UK
[3]King's Policy Institute, King's College London, Virginia Woolf Building, 22 Kingsway, London WC2R 2LA, UK
Full list of author information is available at the end of the article

Background

Estimating the returns from biomedical and health research

Estimating the economic returns arising from health research develops our understanding of how research translates from 'bench to bedside', can be used in advocating the case for future investments in medical research, and demonstrates accountability for public and charitable research funding to taxpayers and donors. Because resources used for publicly and charitably funded medical research, including cancer research, could potentially be put to other purposes for the benefit of society, there is an obligation to demonstrate that such investments represent good value. In the medical field, it is possible to identify illustrative examples of specific research breakthroughs that have contributed to substantial benefit in terms of life-saving interventions, or to major improvements in the quality of life of patients with a chronic disease. However, it is much more difficult to describe systematically the nature and extent of the returns to the investment of a whole body of medical research, some of which may inevitably be less fruitful. Furthermore, there are tensions between advocacy, where interested parties are arguing for more research funding, and more dispassionate analysis, which might conclude that too much money is being spent on research. As noted in an editorial in *Nature* in 2010: 'Most of the attempts to count the economic benefits of investment in science have been derived from the efforts of lobbying groups and funding agencies to justify science spending' [1].

The literature that assesses the value of the benefits of medical research forms a relatively small field in terms of methodology and quality [2,3]. There is a lack of clear consensus about key issues, such as the best methods to use to assess the value of the health gains, and there is also variability in the extent to which studies have included all the important components required for a full analysis of the cost-effectiveness of investing in research. As summarised in Table 1, Mushkin [4], in an early study, used a human capital approach to value health gains from all US biomedical research in terms of the productivity gains from having a healthy workforce [4]. This approach has various weaknesses, which were recognised by Mushkin and others [5], including that it tends to overstate the benefits when lost labour can be replaced by unemployed people or through migration, and it undervalues health gains for groups such as the elderly. Funding First [6] advanced the field by building on a different approach based on estimates of the average willingness of individuals to pay for small reductions in the risk of death. They used this figure to value the increased longevity of the US population. In a background paper for this, Murphy and Topel [7] calculated the enormous economic value that would come from finding a cure for cancer and other diseases, but to date, and using the methods they had adopted, the Funding First report claimed 'the largest returns to investment in medical research have come principally from gains against heart disease and stroke' ([6], page 3).

A broadly similar approach was adopted in a series of Australian studies conducted by Access Economics (2003, 2008 and 2011) [8-10], but expanded to allow for improvements in quality of life based on disability adjusted life years (DALYs). In the 2003 version of the report, no allowance was made for the elapsed time between research (input) and improved health and wellbeing (outcome). In the 2008 and 2011 iterations, this was addressed by projecting potential health and wellbeing gains 40 years into the future. In the 2011 report, the authors focused on estimating a return on investment for five specific diseases, including cancer.

To date, only three studies that we are aware of have examined the economic returns from cancer research. Two of those focused on the costs and benefits of US President Nixon's 'War on Cancer' [11-13]. Litchenberg [11] in 2004 examined the contribution of pharmaceutical innovation to increases in cancer survival rates, by looking at the number of new drugs that had been approved to treat cancer after 1971 (when the War on Cancer was declared), and modelling the impact on cancer mortality rates in the US. He estimated that the increase in approved drugs accounted for about 50 to 60% of the increase in age-adjusted cancer survival rates. Although Litchenberg [11] did not compute a rate of return, he did note that the drug costs to achieve an additional year of life per person diagnosed with cancer were well below estimates for the value of a statistical life. Pertinent to the approach adopted in the current study, he concluded: 'Ideally, we would have measured the effects of new cancer drugs on the number of quality adjusted life years (QALYS), but were unable to do so due to lack of data'. In two related papers, Sun *et al.* [12] and Lakdawalla *et al.* [13] followed a similar conceptual approach in quantifying the value of gains in cancer survival, but directly compared this with the costs of research and development (R&D). They estimated that improvements in cancer survival in the US between 1988 and 2000 created 23 million additional life years, equivalent to roughly US$1.9 trillion of additional social value, implying that the average life year gained was worth US $82,000. As with Litchenberg [11], Sun *et al.* did not calculate a return on investment but noted that 'These calculations suggest that from the patient's point of view, the rate of return to R&D investments against cancer has been substantial.' The third study to look explicitly at cancer is the Deloitte Access Economics study [10] cited above. In that report, the authors looked at the rate of return from current (2000 to 2010) research investment in cancer by the Australian National Health and Medical Research Council (NHMRC) and compared this with

Table 1 Methods used in various studies to assess the benefits from health research

Study/features	Mushkin (1979) [4]	Funding first (2000) [6]	Access economics (2003) [8]	Access economics (2008) [9]	HERG *et al.* (2008) [3]	Access economics (2011) [10]
How health gains were assessed	Top-down by disease category: overall gain in each category not linked to specific intervention. Attributed 20 to 30% of total gain to R&D. Reduced morbidity difficult to assess because little reduction in days off work because of sickness. Adjusted the raw data, for example, by applying historical Army and Navy data as an index to record the decline in sickness.	Top-down: overall gain in mortality not linked to specific interventions. Attributed roughly one-third of the total gain to R&D, plus 'some fraction of the credit for the other two-thirds.'	Top-down: overall gain in mortality and morbidity not linked to specific interventions. Attributed 50% of the total gain to R&D.	Top-down: as in the 2003 study, overall gain in mortality and morbidity not linked to specific interventions. Attributed 50% of the total gain to R&D.	Bottom-up: identified research-based interventions, then quantified health impact.	Top-down: overall gain in mortality and morbidity for five disease areas not linked to specific interventions. Attributed 50% of the total gain to R&D.
How health gains were valued	Human capital approach, that is, values attached to lives saved between one period and the next, based on potential future earnings, plus calculation of value of potential working time no longer lost due to sickness.	Used a comparatively high 'willingness-to-pay' value derived from labour economics.	Used the same comparatively high 'willingness-to-pay' value as Funding First.	Used a higher 'willingness-to-pay' estimate than the 2003 study, this time derived from a meta-analysis of international studies.	Used a comparatively low, but arguably realistic, value of health gain by adopting the figure implied by the current level of NHS spending, that is, the opportunity cost of a QALY within the current NHS budget.	Used a lower 'willingness-to-pay' estimate than that used in the 2008 study, in line with Department of Finance and Deregulation guidance.
Proportion of national health gain allocated to national research	Not discussed as a major issue; we assumed it to be 100%.	Not discussed as a major issue in Funding First; we assumed it to be 100%.	Used proportion of global research conducted in Australia (2.5%) to determine the proportion of the total research-based health gain to attribute to Australian research.	Uses bibliometric analysis-based estimate of Australian share of global research output in clinical medicine (3.04%).	An analysis of citations of UK research on UK clinical guidelines suggests average best estimate of 17% linked to UK research.	Uses an updated bibliometric analysis-based estimate of Australian share of global research output in clinical medicine (3.14%).
Costs of health care considered?	No, at least not as a separate item to net-off against the value of the health gains.	No in initial headline figures, but Yes in later analysis: 'the gain in the value of life, net what was spent to attain the longer life, is just 15 percent smaller.'	No, did not net-off the healthcare costs required to achieve the health gains.	No, did not net-off the healthcare costs required to achieve the health gains.	Yes, did net-off the health care costs required to achieve the health gains.	Did not net-off health care delivery costs, but did consider avoided health system expenditure due to gains in wellbeing.
Considered elapsed time between research and health gains?	Yes: 10 years.	Acknowledged time lags between research and benefits but this was apparently not brought into calculations.	No, compared research expenditure and health benefits in the same year. This implies the health gains from research are instant.	Yes: 40 years, with range of 20 to 60 years used for sensitivity analyses.	Yes: an analysis of citations of UK research on UK clinical guidelines suggested average best estimate of 17 years lag.	Yes: same assumption of 40 years as was used in 2008 study. No sensitivity analysis around elapsed time.
How the overall rate of return calculated	IRR of 47%.	Not brought together to provide an overall IRR.	An overall benefit/costs ratio for health research of 2.40.	An overall benefit/costs ratio for health research of 2.17.	IRR of 9% for CVD research combined with 30% for GDP benefits.	Benefit-cost ratios for five disease areas: 6.1 (CVD); 2.7 (cancer); 1.1 (SIDS); 1.2 (asthma); and 0.7 (muscular dystrophy).

Abbreviations: CVD cardiovascular disease, *GDP* Gross Domestic Product, *IRR* internal rate of return, *NHS* National Health Service, *QALY* quality adjusted life years, *R&D* research and development, *SIDS* Sudden Infant Death Syndrome.

gains in wellbeing using DALYS projected for 2040 to 2050. In doing so, they estimated the net benefit of NHMRC R&D between 2000 and 2010 to be AU$1.96 billion with a cost/benefit ratio of 2.7; that is, for every AU$1 million invested in cancer research they would anticipate a return worth $1.7 million.

A recurring theme in these studies is the extent to which health gains can be attributed to research-inspired medical advances. Funding First and Access Economics adopted a 'top-down' (or macro) approach that took a measure of the overall national health gain from various fields of medicine, and then assumed that a proportion was attributable to medical research. One way of addressing this problem of attribution is by examining in a bottom-up manner the impacts of specific projects or programmes of research by tracing forwards from the research to the benefits that arise. Here, considerable progress has been made using the Payback Framework [14-18], but this has relied on the development of specific resource-intensive case studies. Other studies have made progress in analysing the value of the health gains associated with a series of clinical trials [19], but the major challenge faced by these types of studies is attribution: that is, how to show that the health gains that have arisen can be attributed to specific pieces of research.

In 2008, we published a report, funded by the Wellcome Trust and UK Medical Research Council, which aimed to build on the advances in previous studies and address the existing limitations, so as to develop an approach that could be used to measure the economic benefits accruing from publicly and charitably funded medical research [3]. We analysed two major elements of economic returns: the broad impact on the UK Gross Domestic Product (GDP) and the specific net monetary benefits (NMB), defined as the health benefit valued in monetary terms minus the cost of delivering that health benefit which arose from the UK application of relevant UK research. Our analysis of the existing evidence on the GDP or 'spillover' benefits, based largely on US studies from a number of areas of research and certainly not specific to any particular area of medical research, suggested a best estimate of an internal rate of return (IRR) of around 30%. We estimated the NMB of the health gain using methods similar to those used here, giving an IRR of 9% for cardiovascular research. This meant that a GB£1.00 investment in publicly/charitably funded CVD research produced a combined stream of benefits thereafter, equivalent in value to earning £0.39 per year in perpetuity. (We also estimated the NMB from mental health research, which produced an IRR of 7%; however, this was based on a more limited analysis because of data limitation and uncertainties around the effects of interventions in mental health, which meant that we were less confident in the results than we were for the CVD results).

These estimates of the IRR have been widely used in policy circles in the UK and beyond [20-23], and in the absence of any other estimates of the economic impact of biomedical research, the figures have often been used as proxies of the economic impact of medical research more broadly. A consortium of funders (Wellcome Trust, National Institute of Health Research, Cancer Research UK (CRUK), and the Academy of Medical Sciences) commissioned a study to further validate the approach and to explore whether the IRR from the net value of the health benefits in another area, cancer, was similar or not. Thus, this study aimed to estimate the economic returns from UK publicly and charitably funded cancer research on improved health outcomes in the UK specifically. As with the previous CVD study, we accept that there are international benefits of UK research, but this was not in the scope of the current exercise, although as we note, this is an area that warrants further investigation. In addition, and as reported separately [24], we undertook five exploratory case studies to understand qualitatively the complexity of how research translates into health benefit.

We present the methods used for the four main steps that provided the estimated parameters to enable us to calculate the economic returns from the NMB of the UK health gains that we attributed to past UK publicly and charitably funded cancer-related research, and present the results expressed as estimates of the IRR, with sensitivity analyses to illustrate the effects of some of the key uncertainties. Finally, we explored the significance of our findings in the context of previous studies and the wider policy debate on R&D investments and economic impact; we detailed the limitations of our approach; and we developed a research agenda for this fledgling field.

Methods

Overall conceptual approach

Four key sources of data were needed to estimate the IRR of the NMB of the health gains arising from cancer research:

- a time series of the public and charitable funding of cancer-related research;
- a time series of the NMB of cancer health gains, derived from the monetised health benefits and the healthcare costs for selected interventions[a];
- an estimate of the elapsed time between the investment (research funding) and return (health gain) associated with those interventions; and
- an estimate of the amount of health gain that should be attributed to public and charitable research investment in cancer-related research in the UK.

With these four data inputs, we then calculated a rate of return on the investment in cancer research.

It should be noted that the costs of private sector R&D investments are accounted for in our analysis as elements within the cost of delivering health care, which are netted off in the NMB. The costs to the health service of medical interventions produced by the private sector include the return to the private sector on its R&D investments.

Estimating public and charitable funding of cancer-related research

The leading funders of cancer research in the UK were identified by examining the National Cancer Research Institute (NCRI) Cancer Research Database. Between 2002 and 2011, the top 10 funders consistently accounted for over 95% of cancer research spend by the 21 NCRI partners.[b] Estimates of annual cancer-related research funding between 1970 and 2009 were assembled for these 10 organisations plus an estimated contribution to cover Funding Council support for cancer research (the Higher Education Funding Council for England and similar bodies in Wales, Scotland and Northern Ireland provide a performance-related block grant to UK universities based on the quality and volume of research). A detailed account of how we estimated these 11 time series is provided (see Additional file 1).

As also discussed in detail in Additional file 1, in estimating research spend for the Funding Councils and the Department of Health (DH)/NHS, we had to derive a figure specifically for cancer-related research activity in the UK. We settled on a central estimate of 10% of total publicly and charitably funded health and biomedical research activity, and we also assumed it to be constant over the time period. This estimate was derived from a number of independent sources, as follows

- Medical Research Council (MRC) spending on cancer research averaged 9.8% of their total investment (range: 4.6% to 16.7%) between 1970/1 and 2009/10.
- Wellcome Trust cancer funding was more erratic, ranging between 1%[c] and 38%, with an average of 14.5% of expenditure being on cancer research.
- The proportion of peer-reviewed research papers in oncology as a percentage of all UK biomedical outputs averaged 9.2% (range: 8.5% to 9.5%) between 1988 and 1995 [25].
- The proportion of peer-reviewed research papers in oncology research (as a percentage of all NHS research outputs) was 12% between 1990 and 1997 [26].
- The proportion of mainstream quality-related (QR) funding allocations by the Higher Education Funding Council for England for 'Cancer studies' (that is, Unit of Assessment 02) between 2009 and 2012 was around 6% of the total biomedical allocation (that is, Unit of Assessments 01 to 15 and 44).[d]

Given the importance of this estimate of 10% for the proportion of research activity that is related to cancer (for those sources where we had no actual breakdown), we also looked at the effect of lower and higher estimates of 7.5% and 15%, respectively, in the sensitivity analyses.

Estimating the NMB from cancer-related research

This element of the research required estimates of the lifetime QALYs gained and the net lifetime costs to the NHS of delivering those QALYs for research-based interventions provided in each year of the period 1991 to 2010. The general methods mirrored those used in the 2008 study [3] on the returns on investment in CVD research, and again built up the aggregate net benefits from the bottom up, aggregating the QALYs gained and the net NHS costs from the use of specific interventions. This approach required: 1) identification of the key relevant cancer interventions and their level of usage during the relevant period; and 2) estimates of the QALY gains and NHS costs associated with the interventions. From this information, the NMB was calculated as the health benefit valued in monetary terms (determined by the quantity of health benefit and a decision-maker's willingness to pay for that additional benefit) minus the cost of delivering that health benefit.

In the CVD study, our starting point was previously published research identifying the cardiovascular interventions that had contributed most health gain [27]. No equivalent studies for cancer were identified that could provide a comparable basis for deciding which interventions were, quantitatively, the most important to include in the analysis. Thus, the three main steps for quantifying the total NMB associated with cancer interventions were: 1) to identify the cancer interventions that were the likely major sources of benefits; 2) to identify appropriate estimates of NMB per patient for that subset of cancer interventions; and 3) to construct a time series (for 1991 to 2010) of the number of patients receiving each of these subsets of cancer intervention in the UK.

Identifying the key cancer interventions

At the outset of the study, we had a number of discussions with cancer research experts to provide us with a broad understanding of the main developments in the field over the past 20 years. Informed by these discussions, we quantitatively identified those areas that had resulted in the largest health gain in the UK since 1990, arising from three main sources: 1) key cancers where research and resultant health policies have led to health gains

through a reduction in incidence; 2) key cancers for which screening programmes have led to health gains because of early detection; and 3) key cancers where there have been the most significant health gains from increased survival.

To identify areas where a reduction in incidence has been observed, cancer incidence data in the UK were analysed, using UK incidence rates between 1990 and 2008 [28], to calculate a percentage change over the period. This percentage change was then multiplied by mid-period UK incidence (the average per year for 1999 to 2001 [29]) to estimate an absolute change in incidence. Four cancer types have seen significantly larger reductions in incidence between 1990 and 2008: lung (6,500), stomach (4,400), bladder (4,400), and cervical (1,400) cancers. Additional file 2 gives full details for the 21 cancers. The literature was consulted to identify possible causes for these reductions in incidence. Overwhelmingly, smoking prevention and cessation was cited as the reason for a reduction in lung cancers [30]. Falls in rates of stomach cancer are also thought to be linked to smoking along with declines in *Helicobacter pylori* and improvements in diet [29,31]. The picture is less clear, given changes in the ways these cancers are coded, but bladder cancer has been shown to be associated with smoking too [32], which may account for the decline in rates. The fall in cervical cancer can be largely attributed to the roll-out of cervical screening since the 1980s, which in addition to detecting cancers, is able to pick up pre-cancerous abnormalities and so reduce the incidence of cancer. This has led to a focus on reduction in smoking and on cervical screening.

In addition to cervical screening (which has been in its present form since 1988), there are currently two other national screening programmes in the UK aimed at early detection of cancers: breast cancer screening (introduced in 1988) and colorectal cancer screening (introduced in 2006). There is evidence that all three programmes have reduced mortality [33-35], and should be included in our list of priority interventions.

There have been substantial advances in cancer treatment in recent decades, which have led to valuable health gains. Surgical techniques remain a cornerstone of treatment, aided by ever-refined radiotherapy methods. The advent of new cytotoxic therapies, as well as hormonal and biological therapies, has greatly increased the available treatment options. Given the breadth of these treatments (and backed up by the number of treatments that expert opinion had identified) it was necessary to limit the focus of our estimation to a subset, which we expected to include most of the health gains likely to have been observed between 1991 and 2010. Data on changes in survival were used as a proxy for health gains. Data were compiled for cancer types on 1-year and 5-year survival rates from CRUK [36] and the Office for National Statistics (ONS) [37] (see Additional file 2). Rates were calculated as percentages for the period 1986 to 1990, and compared with those in 2005 to 2009 to calculate a change in the proportion of people surviving 1 and 5 years after diagnosis. This change in rate was then multiplied by the 'mid-point' incidence in 1999 to 2001 to estimate the additional number of people surviving. The same three cancer types (albeit in slightly different order) were found to have the highest number of additional people surviving for both 1 and 5 years; these were prostate, colorectal, and breast cancer. These three accounted for 73% of the estimated gains in 5-year survival. Using clinical guidelines published by the National Institute for Health and Care Excellence (NICE) a set of the main interventions for each of these three cancer types was identified. These interventions were all treatments, because, although there have been improvements in diagnostics and service configuration, it was assumed that the benefits derived from these should, in principle at least, be reflected in the number of people accessing treatment and in measures of treatment effectiveness.

Identifying estimates of per-patient NMB for the set of cancer interventions

As a result of the approach outlined above, estimates of per patient cost and effects were then obtained from published studies for the following prioritised areas:

- Smoking prevention/cessation
- Screening programmes: cervical, breast, and bowel cancer.
- Treatment of: breast, colorectal and prostate cancer.

Smoking prevention/cessation

The area where we adopted a very different approach to that which we had previously used for CVD was smoking. In that study, we restricted the analysis to the costs and benefits arising from NHS smoking cessation interventions. Cancer research has not only unequivocally shown the causal link between smoking (both active and passive) and both cancer and the risks of cancer (and other health problems) but also the effectiveness of various national interventions in reducing smoking rates. This cumulative evidence has contributed to a slow but steady change in smoking behaviour both through direct effects on individual behaviours and through the many non-NHS interventions in the UK (such as legislation and taxation) which have followed from, and been made possible by, this evidence, and have encouraged existing smokers to quit and discouraged others from taking up smoking, as summarised in Figure 1. Therefore, health gains from research include not only the benefit from getting smokers to quit (aided or not by the NHS), but also in preventing non-smokers from ever starting smoking. A recent modelling study for the UK DH Policy Research Programme provided

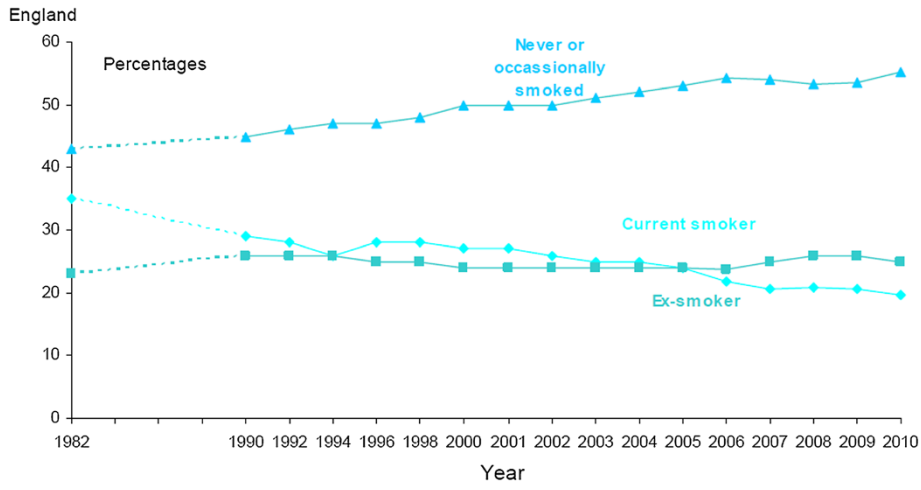


Figure 1 Smoking behaviour in England, 1982 to 2010. Source: General Lifestyle Survey 2010. The Office for National Statistics. Copyright © 2012, re-used with the permission of The Office for National Statistics.

estimates of lifetime life years gained and cost savings to the NHS of non-smokers and ex-smokers compared with smokers [38]. The model accounted for the mortality benefits from not smoking associated with lung cancer, myocardial infarction, stroke, and chronic obstructive pulmonary disease. In the absence of age-specific smoking rates, we used the estimates for men and women aged 35 years, and adjusted these to take account of the proportion of life years gained resulting from lung cancer reduction and also the adjusted life years gained by the population mean utility values for the relevant ages in order to estimate QALYs gained [39].

Screening programmes

To estimate the NMB of each of the three screening programmes, we identified the most appropriate economic evaluations that modelled the lifetime costs and effectiveness of offering the screening programmes as delivered in the UK. For both cervical and bowel cancer screening we used assessments that had informed relevant screening policy decisions [40,41]. In the case of cervical screening, we adjusted the figures presented as life years gained by an appropriate age/sex population utility values to give an estimate of QALYs [39]. For breast cancer, we used a recently published economic evaluation that had used a life-table model to assess the overall cost-effectiveness of the NHS screening programme which based its assessment of effectiveness on the findings of the Independent UK Panel on Breast Cancer Screening, and took account of the uncertainty of associated estimates of benefits, harms, and costs [33,42]. In all three cases, these models used take-up rates that were the same or very similar to those observed in the relevant screening programme during the period in question.

Treatment programmes

The full list of treatment interventions included in the estimation of health gains for each cancer site are shown in Additional file 3. These were determined based on NICE Clinical Guidelines (CG131 for colorectal cancer [43], CG80 and CG81 for breast cancer [44,45] and CG58 for prostate cancer [46]) and cross-checked to ensure that relevant interventions identified by experts were included. Patient sub-groups were recognised where distinction in treatments was made, or where likely differences in cost and benefits existed. In breast cancer, for instance, this distinction was made for node-positive cancers, oestrogen receptor-positive cancers, HER-2-expressing cancers, and pre/post-menopausal incidence of cancers, and between early-stage and late-stage cancer. Historical comparators for each intervention identified from the contemporary guidelines were then identified back to 1991.

For each of the treatment options considered, published economic evaluations were used to estimate per patient costs and benefits (measured as QALYs). Searches were conducted using the NHS Economic Evaluation Database and MEDLINE to identify economic evaluations of prostate, breast, and colorectal cancer interventions. UK-specific estimates were preferred, but international evidence was used where no appropriate UK estimates were available. Where they were available, NICE technology appraisals and National Institute of Health (NIHR) Health Technology Assessments were used as the most relevant sources (see Additional file 3). Where exceptionally non-UK cost-effectiveness data had to be used, costs were converted using purchasing power parity exchange rates.

Constructing a time series (1991 to 2010) of usage of cancer interventions

To estimate total NMB for the period, per-patient QALY gains and net costs for each intervention were multiplied

by the total number of new patients receiving each intervention in each year. We used the following methods to estimate the time series of usage for the selected interventions.

For smoking reduction/cessation we used figures derived from the data on the proportions of smokers, ex-smokers and non-smokers for England for each of the years to estimate the net change per year in QALYs gained and NHS savings achieved, and related these to population data for the UK as a whole [47].

For cervical and breast screening programmes, we used figures for the relevant size of the UK age group in each year to whom screening was first offered (age 25 for cervical and age 50 for breast). For bowel screening we used the numbers first offered screening as the programme began to be rolled out.

To estimate the numbers of people receiving each treatment intervention over time two primary sources were used. For surgical procedures (for example, colorectal excision, liver resection and ablation, prostatectomy, orchiectomy, mastectomy and lumpectomy) Hospital Episodes Statistics [48] were utilised. To estimate the numbers of people receiving drug interventions, data on Net Ingredient Cost (NIC) of drugs to the NHS were used. These data were gathered from Health and Social Care Information Centre (HSCIC) data publications [49], which give details of the total cost of a particular drug prescribed in primary care (for the Prescription Cost Analysis) and secondary care (Hospital Prescriptions Audit Index) in each year. For some drugs, this information was not available for the whole of the time period, in which case assumptions were made on the basis of launch year and the most recent available time point. If the launch year occurred during the period 1991 to 2010, a linear interpolation with launch year at £0 NIC was performed. For drugs that were not launched during the period, a last value carried back approach was adopted, using the most recent year of historical data. From the NIC, the cost and length of a typical regimen (as estimated by NICE costing templates where possible) were used to calculate the number of complete treatments delivered and hence the number of people receiving a particular drug in any given year. This was then proportioned across the indications of a drug and particular patient group (for example, early and late cancers, or multiple cancers).

For some older drug interventions, NIC data were not publicly available for any of the years of interest. In these instances, NICE estimates of the proportion of patients likely to receive interventions (based on guidance costing templates) were combined with data on incidence to estimate usage numbers.

For radiotherapy, there was a paucity of data on usage. Data from the National Clinical Analysis and Specialised Applications Team (NATCANSAT) were available for 2009/10, giving the number of episodes of radiotherapy.[e] It was estimated that 70% of these episodes would be for primary treatment of a cancer. The number of primary radiotherapy episodes was estimated as a proportion of the incidence of each cancer in 2009/10. This proportion was applied historically to incidence in order to estimate radiotherapy treatment.

The component figures of numbers of people receiving treatment interventions were all derived from data for England. To produce a UK estimate (needed because the research spend data is for the UK) figures were adjusted by a factor reflecting England's proportion of the adult UK population. The screening was based directly on relevant UK population data, and for smoking behaviour the time series data were for England, but have been applied to the UK population. All cost estimates were adjusted to 2011/12 prices using the Hospital and Community Health Services Pay and Prices Index [50].

For the calculation of NMB, we used for the base case an opportunity cost value of a QALY as used by NICE in its decision-making [51,52]. This value reflects an estimate of the opportunity cost in terms of QALYs forgone elsewhere in the health service within its fixed budget. Given that public spending on health research can justifiably be seen as a decision to spend on research rather than directly on current healthcare, this opportunity cost value is appropriate to the public decision regarding research funding. In this study, as previously for CVD, we characterised NICE's threshold range as equivalent to an average of £25,000 per QALY, but considered a broader range of values in the sensitivity analysis, including a value of £70,000, which would be broadly consistent with the commonly proposed QALY threshold of 3 times GDP per capita [53].

Analysis of UK clinical guidelines to estimate elapsed time and rate of attribution

In the 2008 report on CVD research, the references cited in a sample of clinical guidelines were analysed to inform the estimate of the elapsed time between research spend and net health gain, and the proportion of net health gain that could be attributed to UK research [3]. In the current study on cancer research, we replicated this approach.

In total, 31 national clinical guidelines, which provided a broad representation of cancer practice in the UK, were identified. Twelve were published by NICE and a further twelve by the Scottish Intercollegiate Guideline Network (SIGN). The remaining seven guidelines were published by either the Royal Colleges or the National Cancer Screening Programme. The reference sections of these guidelines were reviewed: five had no reference list (four published by NICE, one by the National Screening Programme) while one screening guideline had no

references to peer-review journals (that is, it referenced only policy and practice documents). These six guidelines were excluded from our sample. We then used a bespoke computer programme to extract references from the electronic PDF version of each guideline; in three cases the automated reference extraction failed (because papers were not referenced in a recognised format), leaving us with a sample of 22 national guidelines.

Of the 5,627 references cited in the 22 guidelines, 4,416 references (78%) were automatically extracted, excluding duplicate references within a guideline (see Additional file 4 for breakdown by guideline). Nine of these references had no date information and were excluded from the analysis of elapsed time, leaving a total of 4,407 references. The age of a paper cited in a clinical guideline has been termed the 'knowledge cycle time' [54], which is the average difference between the publication date of the clinical guideline and the publication date of the cited papers on the guideline. The knowledge cycle time was calculated for the 22 identified guidelines, and used to inform the estimated elapsed time.

To estimate the rate of attribution to the UK, the 4,416 extracted and de-duplicated references were provided to the Centre for Science and Technology Studies (CWTS) to be matched to their bibliometric database (which is derived from the Web of Science).[f] Of the 4,416 extracted references, CWTS was able to match 4,051 (92%), which formed the dataset to estimate the degree of attribution based on the address field in the cited papers. These addresses were used as a proxy for the location in which the research was conducted, and so it was possible to estimate the proportion of the cited research that was conducted in the UK. The non-matched references included non-serial outputs such as books, journals that are not indexed on the Web of Science, and incorrect references.

Estimation of the rate of return

Using these four key sources of data, we could then attribute a proportion of the estimated total annual NMB of the cancer health gain as being due to UK research, and relate an equal number of years of investment to years of NMB, 'lagged' by an estimate of the average lag between research and benefit. The return was expressed as an IRR, which is effectively the discount rate that would yield a zero net present value. The IRR is convenient in enabling a comparison to be made between non-competing investments of different sizes (as well as providing a direct comparison with our previous study). We recognise the many and various layers of estimates involved. In other circumstances, it might be feasible to express the uncertainty as ranges for each parameter in our overall estimate and undertake a formal probabilistic sensitivity analysis (PSA). However, given the nature of the evidence from multiple sources for the numerous parameters and the necessary judgments involved in drawing together and interpreting the evidence, a comprehensive PSA quantitatively characterising all the uncertainty was not feasible here, and indeed would be liable to suggest a spurious precision. Instead we provide a series of one-way and scenario sensitivity analyses to illustrate the effects of specific variables on the IRR.

Results

Public and charitable funding of UK cancer-related research, 1970 to 2011

Additional file 1 provides our estimated expenditure by year by organisation for the 40-year period, 1970 to 2009, with a summary of cash expenditure provided in Figure 2. Figure 3 illustrates estimated public and charitable expenditure on cancer-related research from 1970 to 2009 in cash and constant 2011/12 prices (the latter for our best estimate). About £15 billion (in 2011/12 prices[g]) of cancer-related research funding was invested during this period. The data presented in Figure 3 are derived from a number of different sources and include various assumptions and estimations. For this reason, we also provided a 'high' and 'low' scenario for total cancer-related research expenditure with a range of £14 to £17 billion. In Figure 3, we also present total public and charitable spending on cancer-related research in cash terms. This emphasises that in real terms (in 2011 prices; the red line) spending fell between 1970 and 1979, then stagnated until 1986, and thereafter increased threefold, from £250 to £850 million, by the end of the time series in 2009.

Net monetary benefit

Table 2 shows the contributions to our total estimates of lifetime QALYs gained from the seven areas we addressed, classified by the year in which the intervention was delivered (or in the case of screening, the year in which those targeted entered the screening programme). Reduction in smoking accounted for 51% of the QALYs gained from the seven areas we prioritised. The other two large sources of QALYs gained were from cervical screening (21%) and breast cancer treatments (19%). The other areas we examined were small contributors by comparison.

Table 3 shows the lifetime net costs to the NHS for each of these areas over the 20-year period. The key points to note here are the high proportion of total net costs accounted for by breast cancer and prostate cancer treatments. Smoking reduction on the other hand reduces net NHS costs, as does colorectal screening, although the latter's introduction late in the period covered means its absolute contribution to reducing overall costs is small.

Table 4 summarises the NMB when the QALYs have been valued at £25,000 and the net costs to the NHS of the intervention and its long-term sequelae have been deducted. It shows how the total NMB (when measured

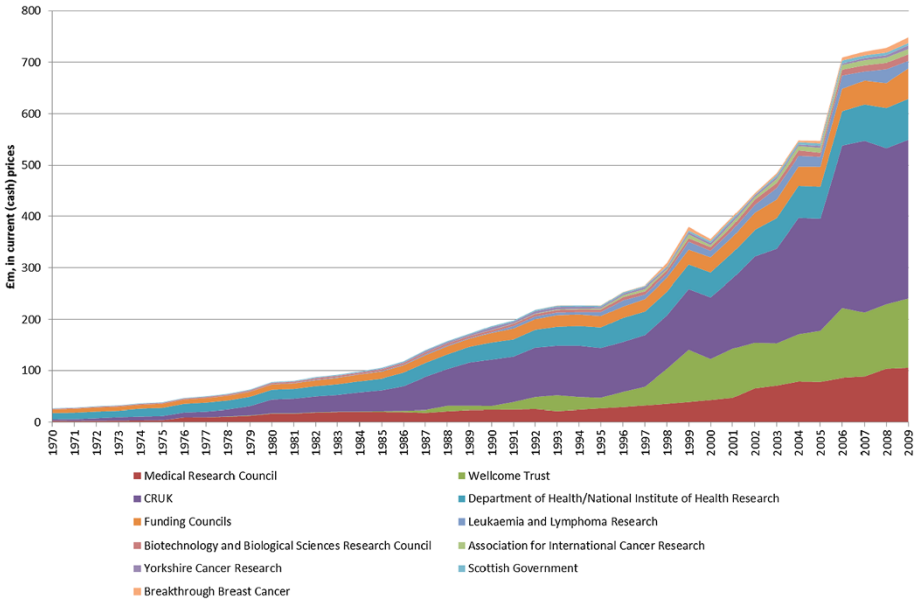


Figure 2 Total estimated public and charitable spend on cancer research by source of funding, 1970 to 2009, in current (cash) prices.

in constant prices) from the research-based interventions that we have assessed has been steadily increasing, with an overall increase of 28% over the 20-year period. Over the whole period, smoking reduction (providing for both QALYs and NHS cost savings) accounted for 65% of NMB, followed by cervical screening (24%) and breast cancer treatments (10%). All seven areas we studied showed a positive NMB when QALYs were valued at £25,000. However, at a QALY value of £20,000, prostate and colorectal treatments and breast cancer screening all showed a negative NMB (that is, their net costs exceeded the valuation of the benefits they provide).

Estimating the elapsed time

The estimate of the elapsed time used in the study was based primarily on the analysis of cited references on clinical guidelines (that is, knowledge cycle time). As illustrated in Figure 4, the mean age of the 4,407 cited

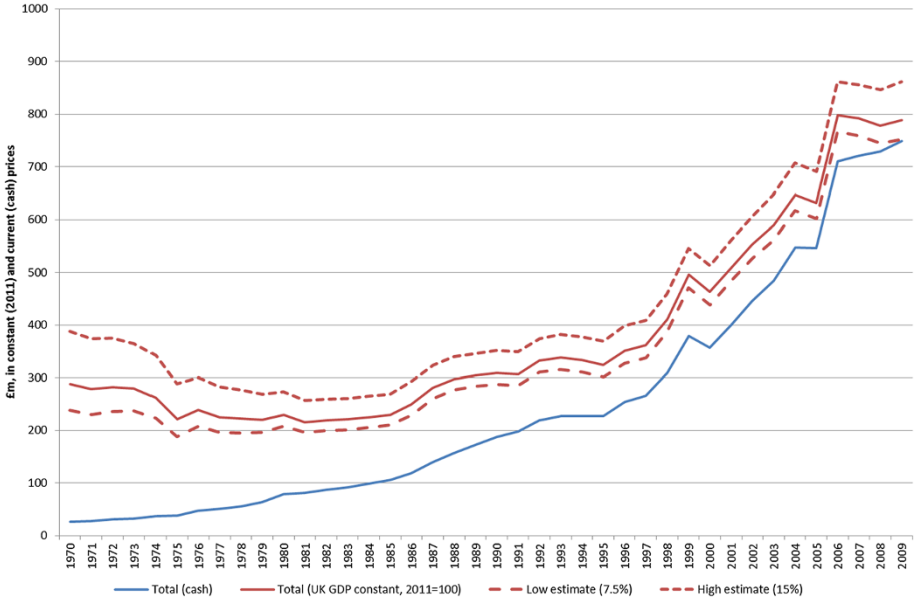


Figure 3 Public and charitable funding of cancer research 1970 to 2009, at constant (2011) prices with low and high estimates, and in current (cash) prices.

Table 2 Contributions of the seven areas to the total estimates of lifetime QALYs gained by year: 1991 to 2010

Year	QALYs (thousands)							
	Treatment			Screening			Smoking reduction	Total
	Prostate cancer	Breast cancer	Colorectal cancer	Cervical cancer	Bowel cancer	Breast cancer		
1991	8	46	6	71	–	2	144	277
1992	9	48	6	70	–	2	144	279
1993	10	46	6	68	–	2	145	276
1994	11	48	6	67	–	2	145	279
1995	10	45	6	65	–	2	145	273
1996	11	46	6	66	–	2	146	276
1997	11	46	6	63	–	3	146	274
1998	11	50	6	59	–	2	147	276
1999	13	53	7	56	–	2	147	279
2000	15	53	7	55	–	2	148	281
2001	17	55	7	54	–	2	149	285
2002	20	56	8	53	–	2	150	290
2003	22	59	8	53	–	2	151	295
2004	24	61	9	56	–	2	152	305
2005	21	62	10	60	–	2	154	309
2006	22	62	12	61	2	2	155	316
2007	23	65	13	60	5	2	157	324
2008	27	68	14	60	7	2	158	337
2009	26	71	15	63	9	2	159	345
2010	25	74	15	65	12	3	161	354
Total	339	1112	173	1225	35	43	3003	5930

Abbreviation: QALY quality adjusted life year.

papers on the 22 guidelines was 8 years, ranging from 0 to 88 years (the median age was 6 years, with an interquartile range of 3 to 10 years). To produce an estimate of elapsed time between spending on research and health gain as required for this study, it was necessary to add on to this value the estimates for the period between the awarding of funding and publication, and the period between recommendation and use. Using the same approach adopted in the 2008 report, we estimated these two periods to total approximately 7 years, giving a best estimated elapsed time between spending on research and health gain of 15 years, with 10 and 20 years arbitrarily selected as lower and higher estimates for sensitivity analyses.

Estimating the amount of health gains that can be attributed to UK research

The estimate of the proportion of the health gain that can be attributed to UK research used in the study was based primarily on the analysis of cited references on clinical guidelines. A total of 4,051 publications were analysed to estimate the proportion of the research that could be attributed to the UK. The overall percentage across all guidelines was 17%, but as shown in Additional file 4, this differed between specific guidelines.

Estimating the IRR from cancer-related research

Our estimates of the NMB produced by year (summarised in Table 4) were then related to the estimated public and charitable spend by year on cancer-related research (summarised in Figure 3) and expressed as an IRR. Calculation of the IRR incorporates our best estimates of the elapsed time of 15 years (low and high estimates of 10 and 20 years) and of the proportion of the NMB that could be attributable to UK research (best estimate 17%: low and high range estimates of 10 and 25%). Thus in our base-case calculation we related 17% of the annual NMB (for each of the 20 years 1991 to 2010) to 20 years of the research investment that had occurred 15 years earlier (that is, for the years 1976 to 1995; in other words, a subset of the 1971 to 2009 series collated). This produced a base-case estimate of the IRR of 10.1%.

As is evident from the methods used, there is inevitably considerable uncertainty around the values of all our estimates. Table 5 presents a series of one-way sensitivity analyses to illustrate the effects of some of the main

Table 3 Contributions of the seven areas to the estimates of lifetime costs to the NHS of services delivered by year: 1991 to 2010

Year	Costs (GB£ million)							
	Treatment			Screening			Smoking reduction	Total
	Prostate cancer	Breast cancer	Colorectal cancer	Cervical cancer	Bowel cancer	Breast cancer		
1991	199	665	181	41	–	34	-277	844
1992	220	687	190	40	–	37	-277	897
1993	241	658	185	39	–	40	-278	887
1994	272	684	185	39	–	42	-278	944
1995	252	646	175	37	–	42	-279	874
1996	278	660	182	38	–	43	-280	921
1997	269	684	181	36	–	53	-281	943
1998	283	720	187	34	–	50	-282	993
1999	336	753	214	32	–	47	-283	1098
2000	391	746	211	32	–	45	-285	1140
2001	456	755	206	31	–	43	-287	1204
2002	519	764	202	30	–	43	-282	1276
2003	571	794	178	30	–	44	-255	1361
2004	614	817	179	32	–	44	-254	1432
2005	545	850	182	34	–	44	-252	1403
2006	572	851	188	35	-5	45	-258	1428
2007	596	881	177	35	-10	46	-252	1473
2008	613	919	181	35	-15	48	-242	1538
2009	606	950	185	36	-20	48	-237	1569
2010	569	986	186	38	-25	56	-240	1569
Total	8403	15469	3755	704	-75	894	-5358	23793

Abbreviation: NHS National Health Service.

areas of uncertainty, and all changes have predictable effects. For NMB, the greatest uncertainty in our calculations probably relates to the magnitude of the benefits from smoking, given the indirect nature of the estimate. Reducing (or increasing) the NMB from smoking by 25% produced an IRR of 8.7% (or 11.2%); for illustration, the (unrealistic) extreme of removing entirely the benefits from smoking from our estimates produced an IRR of 2.4%. The IRR increased as our estimates of research funding were reduced, and the proportion of benefits attributable to UK research increased. It was found to be particularly sensitive to a reduction in the elapsed time. Although taken individually, all of the alternative values we have explored in this sensitivity analysis showed a reasonable rate of return, in combination they could of course have produced a wider range of estimates for the IRR.

Discussion

Taking into account the necessary assumptions made in our approach, the base-case IRR for the NMB from the health gain from cancer research of approximately 10% is remarkably similar to that derived for CVD research, where the IRR derived from the health gain was 9%. These benefits alone provide a return considerably greater than the UK government's minimum threshold of 3.5% for investments, thus suggesting that investment in cancer research is worthwhile. Moreover, given that CVD, cancer, and mental health account for about 45% of the total burden of disease in the UK [55], we might with increasing confidence extrapolate this order of rates of return to the whole of the public and charitable investments in biomedical and health research in the UK. The important caveat to that statement is that the two of the clinical areas that we have analysed in most detail – cancer and CVD –have both benefitted significantly from the changes in smoking over the period analysed.

However, it should be remembered that in our previous study, the rate of return from NMBs of the health gain was less than a third of the rate of return (30%) that we suggested might relate to the broader GDP gains. If we accept that estimate of returns from GDP, then again the

Table 4 Contribution of the seven areas to the estimates of net monetary benefit by year: 1991 to 2010 (QALY value of £25,000)

Year	Net monetary benefit (£ million)							
	Treatment			Screening			Smoking reduction	Total
	Prostate cancer	Breast cancer	Colorectal cancer	Cervical cancer	Bowel cancer	Breast cancer		
1991	8	490	-33	1729	–	7	3885	6085
1992	8	506	-35	1699	–	7	3889	6075
1993	9	485	-33	1659	–	8	3890	6018
1994	10	504	-32	1646	–	8	3894	6030
1995	9	475	-28	1592	–	8	3906	5963
1996	-3	483	-28	1609	–	9	3920	5989
1997	-3	458	-28	1546	–	11	3933	5916
1998	-3	531	-29	1447	–	10	3947	5902
1999	-10	566	-40	1374	–	9	3966	5866
2000	-15	579	-36	1350	–	9	3989	5876
2001	-25	630	-30	1310	–	9	4018	5910
2002	-8	647	-5	1283	–	9	4037	5963
2003	-14	677	24	1294	–	9	4035	6025
2004	-7	705	49	1376	–	9	4064	6195
2005	-15	694	67	1460	–	9	4099	6314
2006	-16	697	106	1484	64	9	4138	6481
2007	-23	736	149	1469	128	9	4167	6635
2008	51	786	171	1474	192	10	4192	6876
2009	46	816	178	1528	256	10	4216	7050
2010	66	854	183	1596	320	11	4253	7282
Total	65	12318	566	29927	960	179	80437	124452

Abbreviation: QALY quality adjusted life year.

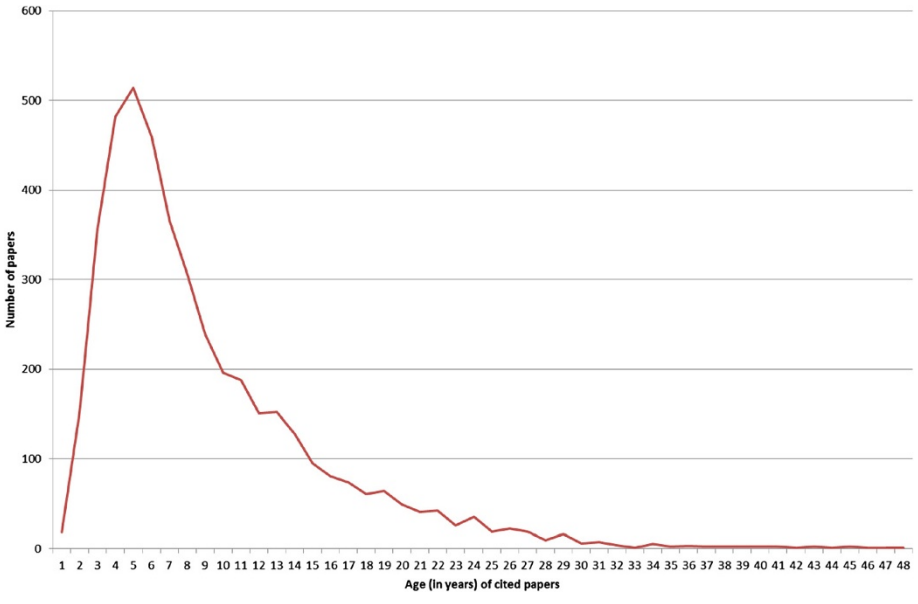


Figure 4 Age of papers cited on cancer clinical guidelines.

Table 5 IRR: one-way sensitivity analyses

Analysis	IRR, %
Base case[a]	10.1
Research funding estimate	
Low	10.8
High	8.7
'Value' of a QALY, GB£	
20,000	8.0
30,000	11.7
50,000	16.1
70,000	18.9
Elapsed time, years	
10	14.6
20	7.4
Attribution to UK research, %	
10	6.1
25	13.0
Effect of smoking cessation	
Decrease NMB by 25%	8.7
Increase NMB by 25%	11.2
Omitting benefit of smoking reduction	2.4

Abbreviations: IRR internal rate of return, *NMB* net monetary benefits, *QALY* quality adjusted life year.
[a]Best estimate for research funding and net monetary benefit, QALY value of £25,000, elapsed time of 15 years, and attribution to UK research of 17%).

overall returns from cancer research would be in the order of 40%. However, as we noted then, although this estimate was based on the best available information, it was generated from a small empirical literature, much of it US-centred and only a proportion specific to medical research. From the papers reviewed, a rate of return between 20% and 67% was identified and we took 30% as our 'best estimate'. The current study did not revisit this aspect of the return to investment. As discussed below, we recommend that future research should aim to update and improve on these estimates.

What this paper contributes

In this study, our main methodological contribution was to further validate the bottom-up approach we developed in the original *Medical Research: What's it Worth?* study [3]. This new application strengthens our argument that the bottom-up approach represents a significant improvement on earlier attempts to estimate economic returns from research, as it attempts to directly attribute health gains (as measured by QALYs) to research-derived interventions. The alternative 'top-down' approaches face the fundamental problem of starting with changes in mortality or morbidity over time, and attributing an estimated proportion of these changes to biomedical and health research. In addition, and in line with our previous work, we have taken into account the costs of delivering the health gain and the elapsed time between research investment and health gain, which earlier studies had largely failed to do.

Key assumptions and caveats

Despite validating and further developing the approach, there are still a number of key assumptions and caveats in our estimate of the economic returns from cancer-related research. Given this, we would be the first to acknowledge that the bottom-up approach by necessity relies on these assumptions, and that our findings do need to be treated with appropriate caution. We document these assumptions in the interests of transparency, and to stimulate further research. The key assumptions are as follows.

- **Our base-case value of a QALY is £25,000.** Obviously, and as demonstrated by our sensitivity analysis, the IRR is sensitive to the assumed value of the health gain measured as QALYs. Our base-case assumption is consistent with our analysis of the returns to CVD research, and reflects the mid-point in the range of values (of £20,000 to £30,000) cited as the normal criteria for acceptance of interventions by NICE [51]. More recently, NICE has increased this threshold, up to around £50,000, for certain treatments that provide end-of-life benefits, particularly late-stage cancer treatments [52]. At the same time it has seemed to re-emphasise that the £20,000 threshold should apply unless there are special circumstances. Although this leaves uncertainty about the most appropriate value here (as reflected in our sensitivity analysis), conceptually the argument remains that this 'opportunity cost' value of a QALY should apply to an assessment of research in that investing in health-related research can be seen as an alternative to spending the money directly on current health care. We note, however, that other studies in the US and Australia have used much higher values, reflecting individual willingness to pay for health gains, and we have illustrated in a sensitivity analysis the effect of using a value of the order of three times GDP per capita [53].
- **The total NMB for interventions not covered is assumed to be zero.** Our IRR calculation assumes that all other cancer treatment developments/ interventions that we have not specifically included have, in aggregate, no effect on the NMB, because for these, the monetised value of the health benefit is equal to the cost of delivering the benefit.
  In reality, there may be some areas that we have not covered for which the NMB is negative because of the high cost of treatment and low incremental

health gain. Conversely, there are may be other areas that generate a significant number of QALYs at a relatively low cost. We are not in a position to know whether the net effect of the interventions we did not examine is positive, negative, or zero.

- **The total net flow of knowledge between disciplines is zero.** We have assumed that the flow of knowledge is the same into and out of different research fields, and from each research field into the cognate treatment areas. However, we know that research is unpredictable and diffuse, and there may be research disciplines that contribute more than they gain from other areas. One could argue that some of the reduction in mortality from diseases other than cancer that arises as a result of the reduction in smoking (e.g. CVD) which we have excluded, should in fact be included as having been achieved as an additional advantage arising from the evidence of the effect of smoking on lung cancer.
- **All health gain from treatments is captured in the estimates of the health gain from specific interventions.** We have assumed that in principle the health gain from improved service configuration and all other supportive service changes (including diagnostics and imaging) should be captured in the estimates of the gains from specific interventions. In practice, our estimates of QALY gains are mainly derived from UK-relevant health technology assessments that are extrapolated from trial data, which may provide an imperfect estimate of the gain when the interventions are used in routine NHS practice.
- **The definitions of the cancer-related research used by the research funders captures basic research that may have contributed to developments in this area.** This is clearly the case for the cancer-specific funders such as CRUK, as we included all the research they funded. For MRC funding, we relied on the funder classification which, as discussed in Additional file 1, was broad and thus should include basic research. For the Wellcome Trust, which accounts for around 10% of total cancer funding, we had to rely on search terms. We scanned the list of grant titles selected through this search strategy, and this list suggests that fundamental research is being included, although we cannot guarantee that it all is in fact included. For the remaining two funders – the Funding Councils and the DH/NHS – this would not be an issue, as their time series were derived through an estimate of cancer research activity.
- **The knowledge cycle time and attribution rate were largely determined through bibliometric analysis of clinical guidelines.** As part of this study, and reported separately, we undertook a series of case studies that qualitatively explored how research translates into health benefit [24]. This work demonstrates the complexity of biomedical and health innovation, especially when trying to measure the time it takes for research to develop into health benefits. Although the bibliometric approach provides us with an empirical estimate of both the elapsed time and the rate by which we can attribute UK research to UK health gain, it inevitably is a gross simplification of a complex process.
- **We have made various assumptions about the baseline treatment against which we were looking in research-based developments.** For example, in estimating the net health gain from breast cancer treatments, we did not include benefits from standard mastectomy but just estimated the benefits from subsequent developments.
- **There is a risk that we may have double-counted the NMB for individuals who are treated as a result of screening.** Conceptually, the benefits of screening include the downstream NMB of treatments that result from the screening. However, a number of issues minimise the likelihood of our double counting. First, we did not include (in the treatment calculation) all the benefits of treating an individual disease (for example, breast cancer) but only the additional benefits of improved (research-based) treatments, so any additional people who get 'basic treatment' as a result of screening were counted only as an advantage to screening. Second, the benefits and the future treatment costs of a woman entering a screening programme (which is when we estimated the future QALYs and present value of associated net costs) occur in a future year, often many years ahead, so in taking a 20 year period, there is limited scope for counting both. If we had perfect data and were looking at all treatment benefits over a much longer period, we could in principle look only at the benefits of treatments that would encapsulate all the QALY benefits of screening.

In acknowledging these assumptions, we should make the important point that an underlying principle we adopted throughout this study and our previous work on CVD was to err on the side of caution: that is, to make assumptions that would lead, other things being equal, to a lower rate of return. However, compared with our earlier study of CVD, we are less confident that we have always managed to adhere to the principle of conservatism. For example, as discussed above, there is an implicit assumption in ascribing the IRR to the whole of cancer that everything we have not specifically included has, in aggregate, no effect on the NMB (the value of the

health gain is equal to the costs of delivering it). In reality, the aggregate effect of what we have not considered could be positive, negative, or zero. Another issue is that in the CVD study, conservatism often came from adopting the lower of two (or more) published estimates for specific parameters, but for cancer interventions, we rarely had a choice of relevant data estimates, as discussed in more depth below.

In addition to these specific assumptions, there are a number of other broader issues that add to the uncertainty of our estimates and need to be highlighted.

- **We have evidence of linkage between research and health gains but no formal evidence of causality.** Our analysis relied on the reasonable assumption that these health benefits would not have occurred without the evidence from medical research, and we have illustrated the often complex nature of those linkages in case studies [24]. At one level we have addressed this issue of causality by our bottom-up approach, adding together the benefits demonstrated through clinical trials of new interventions. For these, causality from worldwide medical research is all but a truism. However, even for these, we had to assume that a proportion of the benefit (based on the UK contribution to publications cited in guidelines) arose from UK research. It is possible that some or even all of these interventions might have come into use in the UK even if there had been no UK cancer research, but it is improbable that the same level and timing of benefits would have arisen. Causality could be argued to be less direct for the benefits of the reduction in smoking, which made the largest contribution to the total NMB. It is possible, but implausible, that changes in smoking behaviour might have arisen in the absence of any evidence of the health effects. Certainly, our case studies [24] show that there was an extended lag between the initial evidence of harms to smokers and changes in behaviour, and the UK government probably needed the cumulative evidence that has emerged over several decades, and in particular the evidence of the harms of environmental tobacco smoke, to make the legislative changes in the face of very considerable resistance. There are also additional uncertainties around the magnitudes of NMB from smoking. Of the total £124 billion total NMB, £80 billion (or 65%) arose from reductions in smoking, and the numbers for the increased proportion of the population who were non-smokers or ex-smokers is based on self-reported survey data. In the sensitivity analysis (Table 5), if the NMB from smoking reduction was decreased or increased by (an arbitrary) 25%, the IRR would reduce to 8.7% or increase to 11.2% respectively. Omitting the benefits from smoking reduction entirely reduces the IRR to 2.4%. However, it should be stressed that we estimated only the mortality effects on lung cancer and excluded effects on other cancers (and other disease areas) from smoking, all of which would mean we probably underestimated the impact of smoking reduction. However, taking a perspective of NHS costs only, we have not included costs to other parts of the economy from the various measures to reduce smoking [56].
- **Variable quality of data on the effectiveness of screening.** The three national screening programmes are important elements in our estimates. The clinical and cost-effectiveness evidence for bowel cancer screening is high-quality and trial-based, but the evidence for cervical screening, and even more so for breast cancer screening, is less robust. The recent review [33] of the clinical evidence has provided some clarity to the contentious issue of the net benefits of breast screening, and underpins the relatively simple economic model that we used as the basis of our estimate of NMB, but there is considerable uncertainty around these estimates.
- **There is a lack of robust clinical effectiveness and cost-effectiveness data for some interventions, especially for longstanding treatments.** This was a general problem with well-established surgical techniques (for example, total mesorectal excision, for which no cost-effectiveness evidence could be found) and similarly for some of the hormonal therapies (for example, tamoxifen and goserelin).
- **There are a large number of areas of cancer that we did not consider in our analysis.** Our analysis was based on a prioritised list of cancer types generated from both expert opinion and epidemiological data. By necessity, this meant we did not look at a number of areas (and as noted above, assumed the NMB arising from these areas to be zero).
- **Elapsed time was an important variable in determining the IRR, but one that is conceptually difficult to measure [24].** We wanted to measure the time between research investment and health gain, but neither of these events occurs at one defined point. Research investment may occur over a period, although in many cases, given a typical pattern of investment starting with pilot trials, and building to larger-scale studies and finally randomised controlled trials, the bulk of the research investment may come late in the overall investment period. The point at which the bulk of the health gain occurs is even more difficult to define, and will depend on a

range of factors, such as the type of intervention and the way in which it is implemented. The issue of time lags was identified in the original 2008 report, which suggested that further research is needed.

Given these various issues and the nature of the exercise, which relies on data and estimates from a wide variety of sources, it is not possible to characterise in any formal way the overall uncertainty in our estimates. The sensitivity analysis illustrates the effect on the IRR of alternative values for some of the key parameters, and shows that the broad order of magnitude of the IRR is relatively insensitive to fairly substantial degrees of uncertainty on specific elements of the analysis of what has happened in the past. Moreover, even without this uncertainty, we need to interpret our analysis of what has happened in the past with caution, as follows.

- **Past performance is not an indicator of future performance.** The IRR is based on past performance, and cannot be a guarantee of future returns, particularly for increased levels of research spending. This means that research advocates need to use the estimates provided in this paper very cautiously if wishing to extrapolate them as indicators of likely future returns from research expenditure. Given the near doubling in cancer-related research funding since the turn of the century (Figure 3), there will need to be a similar increase in NMB in the coming decade to maintain the current returns. It is worth noting that the NMB of bowel screening is not fully reflected in the IRR because this screening is of recent introduction, so there is additional benefit that will be realised in the future. Likewise, pharmaceutical interventions are typically priced to maximise the value of the benefit at time of introduction, so the NMB is close to zero. During the coming decade, some of the expensive drugs will come off patent and may be available more cheaply, thus contributing to an increase in the NMB; however, other new and expensive 'on patent' drugs may well be used in preference.
- **We estimated average returns from cancer research, not the marginal returns.** From this analysis, we are not able to say whether the rate of return would have been different if research spending had been higher or lower, and whether at the margin the returns to research investment are increasing or diminishing.
- **The analysis should not be used to make comparative assessments about the value of research into particular interventions/cancers.** Our approach examined a portfolio of interventions/cancer types and we would caution that the detailed data may not be sufficiently robust to make comparisons between interventions within specific cancers.

Future research requirements

Based on the key assumptions, uncertainties and caveats described above, further research is needed in the following areas.

- **A deeper understanding of the international flows of knowledge.** In our model, we estimated the extent to which UK research influences UK practice, using citations on clinical guidelines, and this figure was used in estimating the IRR. However, there is a need for a more nuanced understanding of these knowledge flows and their impact on international health gains; for example, UK research is contributing to health gains beyond the UK. As a result, our current figure underestimates the global value of UK R&D. A study that aimed to measure the health gains, net of healthcare costs, in the rest of the world as a result of UK medical research would address this. At a European level, it would also be interesting to explore how the investments of different European countries in biomedical and health research leads to health gains in other European countries, thereby reinforcing the notion of European solidarity.
- **An improved estimate of spillover effects for UK biomedical and health research.** Public and charitable biomedical and health research expenditure not only leads to health gains, but also makes an important contribution to the national economy. Much of the evidence base for estimating a spillover effect of 30% comes from studies undertaken in the 1960s and 1970s, and/or relates specifically to agriculture research. More recent analyses for medical research are largely based on US data. Furthermore, in this study, we also assumed that the spillovers are independent of disease area but we have no empirical evidence to support whether that assumption is justified or not. Future research should aim to provide empirical estimates of the effects of biomedical and health research for the UK economy, ideally at a disease-specific level.
- **Examine another disease area or time period in which smoking reduction is likely to have a minimal impact.** As illustrated in Table 5, the IRR for cancer research is very dependent on the effect of smoking reduction. It would be valuable to undertake an investigation in another clinical area in which smoking is not important to see whether similar rates of return are found.

Conclusion

It is challenging to move beyond the identification of the benefits from specific examples of research funding and attempt to meet the increasing demands for accountability by systematically measuring returns to the investment of a whole body of medical research. In this paper, we have estimated the economic benefit of public and charitable funding of cancer-related research, and further validated the methodological approach that we originally used in assessing the returns from CVD research. Expressed in 2011/12 prices, total expenditure on cancer-related research from 1970 to 2009 was £15 billion. Over the period 1991 to 2010, the interventions we prioritised in our study produced 5.9 million QALYs and a NMB of £124 billion, allowing for the net NHS costs resulting from them, and valuing a QALY at £25,000. The proportion of the benefit attributable to UK research was 17%. The lag between research funding and impact for cancer treatments was 15 years. Our best estimate of the health-gain IRR from UK cancer-related research was 10%, very similar to that of 9% for CVD research. The results suggest that, despite the uncertainties around the methods and estimates, the historical returns in terms of the NMB of the health gains derived in the UK from public and charitably funded biomedical and health research are substantial, and could by themselves justify the investment made.

Endnotes

[a]We have used the term 'interventions' broadly throughout this paper to include treatments, screening programmes and a wide range of policies and information that have led to changes in smoking.

[b]NCRI members must have an annual cancer research spend in the UK in excess of £1 million, and have an appropriate peer-review system for ensuring the scientific quality of the research that they fund [57].

[c]Up until 2010, the Wellcome Trust had a policy not to fund cancer research. It changed its policy in recognition that the basic research it funded was increasingly having implications for our understanding of cancer.

[d]Data provided by the Higher Education Funding Council for England in personal correspondence.

[e]Data provided at our request by NATCANSAT produced from the national radiotherapy dataset for years 2009 to 2013.

[f]CWTS maintains a bibliometric database of all scientific publications (including health and biomedical research) for the period 1981 to 2013. This dataset is based on the journals and serials processed for the Internet versions of the Science Citation Index Expanded and associated citation indices, the Social Sciences Citation Index, and the Arts and Humanities Citation Index. This database is operated for bibliometric purposes in service contracts under a License Agreement with Thomson Reuters. See [58] for more information.

[g]We used HMG GDP Deflator [59] to estimate constant prices for 2011 (accessed 9 January 2013). We also compared the Biomedical Research and Development Price Index published by the National Institutes for Health Office of Budget ([60]; accessed 9 January 2013), and concluded that there was no material difference for the purpose of the current analysis.

Additional files

Additional file 1: Funding data.

Additional file 2: Incidence and survival data.

Additional file 3: Summary of interventions, their comparators, and sources, and health gain data.

Additional file 4: Guidelines data.

Abbreviations

CRUK: Cancer Research UK; CVD: cardiovascular disease; CWTS: Centre for Science and Technology Studies; DALY: Disability adjusted life year; DH: Department of Health; GDP: Gross Domestic Product; HSCIC: Health and Social Care Information Centre; IRR: the internal rate of return; MRC: Medical Research Council; NATCANSAT: National Clinical Analysis and Specialised Applications Team; NCRI: National Cancer Research Institute; NHMRC: National Health and Medical Research Council; NHS: National Health Service; NIC: Net Ingredient Cost; NICE: National Institute for Health and Clinical Excellence; NIHR: National Institute of Health Research; NMB: Net Monetary Benefit; ONS: Office of National Statistics; QALY: Quality adjusted life year; R&D: Research and Development; SIGN: Scottish Intercollegiate Guideline Network.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

The project was conceived by JG and MB, and designed and executed by all the authors: JG led on the funding analysis; JG, AP, SG on the guidelines analysis; and MG and MB on the assessment of health gain and the calculation of the IRR. All authors were involved in synthesising and interpreting the results. All authors contributed drafts for various parts of the paper, critically reviewing various iterations and approving the final draft submitted.

Acknowledgements

We are grateful for financial support from the Wellcome Trust, Cancer Research UK, the National Institute of Health Research, and the Academy of Medical Sciences. The funders had no influence on the study design, data collection or analysis, decision to publish, or preparation of the manuscript. Through membership of the Steering Group, the funders did provide ongoing suggestions and advice throughout the study, and reviewed the manuscript. We thank the members of the Steering Group for their advice and support, particularly Liz Allen for chairing, along with David Cox, Aoife Regan, Emma Greenwood and Rachel Quinn; Rodrigo Costas and Thed van Leeuwen (CWTS, Leiden) for providing bibliometric analysis; Thomas White at the NCRI for providing us with funding data and reviewing our analysis; and colleagues in our respective organisations for their help in a variety of ways including Steven Wooding for reviewing an earlier draft. We especially extend our thanks to all the cancer experts that we consulted at various stages of the project, particularly: Peter Selby, Mike Richards, Stan Kaye, Georgios Lyratzopoulos, David Dearnaley, Galina Velikova, and Tim Maughan. Needless to say, we are responsible for any errors, misrepresentations, or inaccuracies.

Author details
[1]Health Economics Research Group, Brunel University, Uxbridge, UK. [2]RAND Europe, Westbrook Centre, Milton Road, Cambridge CB4 1YG, UK. [3]King's Policy Institute, King's College London, Virginia Woolf Building, 22 Kingsway, London WC2R 2LA, UK.

Received: 6 March 2014 Accepted: 6 May 2014
Published: 16 Jun 2014

References

1. Editorial: **Unknown quantities.** *Nature* 2010, **465:**665–666.
2. Buxton M, Hanney S, Jones T: **Estimating the economic value to societies of the impact of health research: a critical review.** *Bull World Health Organ* 2004, **82:**733–739.
3. Health Economics Research Group, Office of Health Economics, RAND Europe: *Medical Research: What's it Worth? Estimating the economic benefits from medical research in the UK.* London: UK Evaluation Forum; 2008 [http://www.wellcome.ac.uk/stellent/groups/corporatesite/@sitestudioobjects/documents/web_document/wtx052110.pdf, Accessed: January 2014]
4. Mushkin S: *Biomedical Research: Costs and Benefits.* Cambridge, MA: Ballinger; 1979.
5. Drummond MF, Davies LM, Ferris FL: **Assessing the costs and benefits of medical research: the diabetic retinopathy study.** *Soc Sci Med* 1992, **34:**973–981.
6. Funding First: *Exceptional Returns: The Economic Value of America's Investment in Medical Research.* New York: Lasker Foundation; 2000 [http://www.laskerfoundation.org/media/pdf/exceptional.pdf, Accessed: January 2014]
7. Murphy KM, Topel R: *The economic value of medical research.* New York: Lasker Foundation; 1998 [http://www.laskerfoundation.org/pdf/economicvalue.pdf, Accessed: January 2104]
8. Access Economics: *Exceptional Returns: The Value of Investing in Health R&D in Australia.* Canberra: Access Economics; 2003 [http://www.asmr.org.au/Except.pdf, Accessed: January 2014]
9. Access Economics: *Exceptional Returns: The Value of Investing in Health R&D in Australia II.* Canberra: Access Economics; 2008 [http://www.asmr.org.au/ExceptII08.pdf, Accessed: January 2014]
10. Deloitte Access Economics: *Returns on NHMRC funded Research and Development.* Canberra: Deloitte Access Economics; 2011 [http://www.asmr.org.au/NHMRCReturns.pdf, Accessed: January 2014]
11. Litchenberg FR: **The expanding pharmaceutical arsenal in the war on cancer.** *NBER Working Paper Series* 2004. Paper No. w10328.
12. Sun EC, Jena AP, Lakdawalla DN, Reyes CM, Philipson TJ, Goldman DP: **An economic evaluation of the war on cancer.** *NBER Working Paper Series* 2009. Paper No. 15574.
13. Lakdawalla DN, Sun EC, Jena AB, Reyes CM, Goldman DP, Philipson TJ: **An economic evaluation of the war on cancer.** *J Health Econ* 2010, **29:**333–346.
14. Buxton M, Hanney S: **How can payback from health services research be assessed?** *J Health Serv Res Policy* 1996, **1:**35–43.
15. Hanney S, Mugford M, Grant J, Buxton M: **Assessing the benefits of health research: lessons from research into the use of antenatal corticosteroids for the prevention of neonatal respiratory distress syndrome.** *Soc Sci Med* 2005, **60:**937–947.
16. Wooding S, Hanney S, Buxton M, Grant J: **Payback arising from research funding: evaluation of the Arthritis Research Campaign.** *Rheumatology* 2005, **44:**1145–1156.
17. Wooding S, Hanney S, Pollitt A, Buxton M, Grant J: *Project Retrosight: Understanding the Returns from Cardiovascular and Stroke Research.* Cambridge: RAND Europe; 2011 [http://www.rand.org/pubs/monographs/MG1079.html, Accessed: January 2014]
18. Wooding S, Pollitt A, Castle-Clarke S, Cochrane G, Diepeveen S, Guthrie S, Hortvitz-Lennon M, Larivière V, Ni Chonaill S, O'Brien C, Olmsted S, Schultz D, Winpenny E, Pincus HA, Grant J: *Mental Health Retrosight. Understanding the Returns from Research. (lessons from schizophrenia). Policy Report.* Cambridge: RAND Europe; 2013. [http://www.rand.org/randeurope/research/projects/mental-health-retrosight.html, Accessed: January 2014]
19. Johnston SC, Rootenberg JD, Katrak S, Smith WS, Elkins JS: **Effect of a US National Institutes of Health programme of clinical trials on public health and costs.** *Lancet* 2006, **367:**1319–1327.
20. Cancer Research UK: *Medical Research and the Spending Review*; 2010 [http://info.cancerresearchuk.org/prod_consump/groups/cr_common/@nre/@pol/documents/generalcontent/cr_048533.pdf, Accessed: 20 May 2012]
21. Association of Medical Research Charities: *Written Submission to the Comprehensive Spending Review*; 2010 [http://www.amrc.org.uk/publications/submission-department-business-innovation-and-skills-part-comprehensive-spending-review. Accessed: January 2014]
22. Academy of Medical Sciences: *Initial Response to the Spending Review Consultation*; 2010 [http://www.acmedsci.ac.uk/download.php?file=/images/publication/128393883472.pdf, Accessed: May 2012]
23. European Medical Research Councils: *White Paper II: A Stronger Biomedical Research for a Better European Future.* Strasbourg: European Science Foundation; 2011.
24. Guthrie S, Grant J, Hanney S, Pollitt A: *Investigating time lags and attribution in the translation of cancer research: A case study approach.* Cambridge: RAND Europe; 2014.
25. Dawson G, Lucocq B, Cottrell R, Lewison G: *Mapping the Landscape: National Biomedical Research Outputs 1988–95.* London: Wellcome Trust; 1998.
26. Wellcome Trust: *Putting NHS Research on the Map: An Analysis of Scientific Publications in England, 1990–97.* London: Wellcome Trust; 2001.
27. Capewell S, Critchley J, Unal B: *IMPACT: a Validated, Comprehensive Coronary Heart Disease Model. Overview and Technical Appendices.* Liverpool: University of Liverpool; 2007.
28. **CRUK statistics.** [http://www.cancerresearchuk.org/, Accessed December 2013]
29. **ONS statistics, MB1 series no 32.** [www.ons.gov.uk, Accessed December 2013]
30. Karim-Kos HE, de Vries E, Soerjomataram I, Lemmens V, Siesling S, Coebergh JWW: **Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s.** *Eur J Cancer* 2008, **44:**1345–1389.
31. **National Cancer Intelligence Network.** *Incidence Stomach Cancer Engl* 1998–2007, [http://www.ncin.org.uk/publications/data_briefings/incidence_of_stomach_cancer_in_england Accessed: February 2014]
32. Bine Kjøller B, Raaschou-Nielsen O, Sørensen M, Frederiksen K, Christensen J, Tjønneland A, Overvad K, Chapelon FC, Nagel G, Chang-Claude J, Bergmann MM, Boeing H, Trichopoulos D, Trichopoulou A, Oikonomou E, Berrino F, Palli D, Tumino R, Vineis P, Panico S, Peeters PHM, Bueno-de-Mesquita HB, Kiemeney L, Gram IT, Braaten T, Lund E, Gonzalez CA, Berglund G, Allen N, Roddam A, *et al*: **Tobacco smoke and bladder cancer—in the European Prospective Investigation into Cancer and Nutrition.** *Int J Cancer* 2006, **119:**2412–2416.
33. Independent UK Panel on Breast Cancer Screening: **The benefits and harms of breast cancer screening: an independent review.** *Lancet* 2012, **380:**1778–1786.
34. Libby G, Brewster DH, McClements PL, Carey FA, Black RJ, Birrell J, Fraser CG, Steele RJC: **The impact of population-based faecal occult blood test screening on colorectal cancer mortality: a matched cohort study.** *Br J Cancer* 2012, **107:**255–259.
35. Canfell K, Freddy S, Beral V: **Cervical cancer in Australia and the United Kingdom: comparison of screening policy and uptake, and cancer incidence and mortality.** *Med J Aust* 2006, **185:**482.
36. **CancerStats – Survival.** [http://publications.cancerresearchuk.org/downloads/Product/CS_CS_SURVIVAL.pdf, Accessed: December 2013]
37. **Cancer survival in England - patients diagnosed 2005–2009 and followed up to 2010.** [http://www.ons.gov.uk/ons/rel/cancer-unit/cancer-survival/2005-2009--followed-up-to-2010/stb-cancer-survival--2005-09-and-followed-up-to-2010.html. Accessed December 2013]
38. Godfrey C, Shezad A, Parrott S, Pickett K: *Economic model of adult smoking related costs and consequences for England, October 2010 (Revised April 2011).* York: Public Health Research Consortium Project Report; 2011 [http://phrc.lshtm.ac.uk/papers/PHRC_A4-06_Final_Report.pdf]
39. Kind P, Hardman G, Macran S: *UK population norms for EQ-5D. Discussion paper 182.* York: University of York; 1999. [http://www.york.ac.uk/media/che/documents/papers/discussionpapers/CHE%20Discussion%20Paper%20172.pdf, Accessed: January 2014]
40. Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N: **Liquid-based cytology in cervical screening, an updated rapid and systematic review and economic analysis.** *Health Technol Assess* 2004, **8:**1–78. Table 21.
41. Whyte S, Chilcott J, Cooper K, Essat M, Stevens J, Wong R, Kalita N: **Reappraisal of the options for colorectal cancer screening. Full report for**

the NHS Bowel Cancer Screening Programme. Sheffield: ScHARR, University of Sheffield; 2011. Table 4.2.1.
42. Pharoah PD, Sewell B, Fitzsimmons D, Bennett HS, Pashayan N: **Cost effectiveness of the NHS breast screening programme: life table model.** *BMJ* 2013, **346:**f2618.
43. National Institute for Health and Clinical Excellence. Colorectal Cancer: *The Diagnosis and Management of Colorectal Cancer. CG 131.* ; 2011 [http://guidance.nice.org.uk/CG131, Accessed: December 2013]
44. National Institute for Health and Clinical Excellence: *Early and Locally Advanced Breast Cancer: Diagnosis and Treatment. CG 80*; 2009 [http://guidance.nice.org.uk/CG80, Accessed: December 2013]
45. National Institute for Health and Clinical Excellence: *Advanced Breast Cancer: Diagnosis and Treatment. CG 81*; 2009 [http://publications.nice.org.uk/advanced-breast-cancer-cg81, Accessed: December 2013]
46. National Institute for Health and Clinical Excellence: *Prostate cancer: Diagnosis and Treatment. CG58*; 2008 [http://publications.nice.org.uk/prostate-cancer-cg58, Accessed: December 2013]
47. **Health and Social Care Information Centre.** England: Statistics on Smoking; 2012. Table 2.2 [http://www.hscic.gov.uk/catalogue/PUB07019/smok-eng-2012-rep.pdf, Accessed: February 2014]
48. **Hospital Episodes Statistics.** [http://www.hscic.gov.uk/hes, Accessed: January 2014]
49. **Health and Social Care information Centre.** [http://www.hscic.gov.uk/, Accessed: January 2014]
50. **PSSRU, Unit Costs of Health and Social Care 2012.** [http://www.pssru.ac.uk/project-pages/unit-costs/2012/, Accessed: January 2014]
51. **Guide to the methods of technology appraisal 2008.** [http://www.nice.org.uk/aboutnice/howwework/devnicetech/GuideToMethodsTechnologyAppraisal2008.jsp, Accessed February 2014]
52. **Guide to the methods of technology appraisal 2013.** [http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9, Accessed: January 2014]
53. World Health Organisation: *Macroeconomics and Health: Investing in Health for Economic Development. Report of the Commission on Macroeconomics for Health*; 2001 [http://whqlibdoc.who.int/publications/2001/924154550x.pdf, Accessed: May 2014]
54. Grant J, Cottrell R, Cluzeau F, Fawcett G: **Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study.** *BMJ* 2000, **320:**1107–1111.
55. Murray CJL, Richards MA, Newton JN, Fenton KA, Anderson HR, Atkinson C, Bennett D, Bernabé E, Blencowe H, Bourne R, Braithwaite T, Brayne C, Bruce NG, Brugha TS, Burney P, Dherani M, Dolk H, Edmond K, Ezzati M, Flaxman AD, Fleming TD, Freedman G, Gunnell D, Hay RJ, Hutchings SJ, Ohno SL, Lozano R, Lyons RA, Marcenes W, Naghavi M, *et al*: **UK health performance: findings of the Global Burden of Disease Study 2010.** *Lancet* 2013, **381:**997–1020.
56. Department of Health: *An Audit of the Impact of the Department of Health's Regulations upon Businesses. Final Report.* September 2013. [https://www.gov.uk/government/publications/department-of-health-regulations-impact-on-business. Accessed: January 2014]
57. **The National Cancer Research Institute.** [http://www.ncri.org.uk]
58. **The Centre for Science and Technology Studies.** [http://www.socialsciences.leiden.edu/cwts/]
59. **Her Majesties Treasury GDP data.** [http://www.hm-treasury.gov.uk/data_gdp_fig.htm]
60. **The National Institute of Health price index.** [http://officeofbudget.od.nih.gov/gbiPriceIndexes.html]

10.1186/1741-7015-12-99
Cite this article as: Glover *et al.*: Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. *BMC Medicine* 2014, 12:99

APPENDIX G: Paper 7

Paper 6: Estimating returns of MSK research

Glover M, Montague E, Pollitt A, et al. Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes. Heal Res Policy Syst 2018 161. 2018;16(1):1-24. https://doi.org/10.1186/s12961-017-0276-7

Glover et al. Health Research Policy and Systems (2018) 16:1
DOI 10.1186/s12961-017-0276-7

Health Research Policy and Systems

RESEARCH Open Access

Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes

Matthew Glover[1], Erin Montague[2], Alexandra Pollitt[2], Susan Guthrie[3], Stephen Hanney[1], Martin Buxton[1] and Jonathan Grant[2*]

Abstract

Background: Building on an approach applied to cardiovascular and cancer research, we estimated the economic returns from United Kingdom public- and charitable-funded musculoskeletal disease (MSD) research that arise from the net value of the improved health outcomes in the United Kingdom.

Methods: To calculate the economic returns from MSD-related research in the United Kingdom, we estimated (1) the public and charitable expenditure on MSD-related research in the United Kingdom between 1970 and 2013; (2) the net monetary benefit (NMB), derived from the health benefit in quality adjusted life years (QALYs) valued in monetary terms (using a base-case value of a QALY of £25,000) minus the cost of delivering that benefit, for a prioritised list of interventions from 1994 to 2013; (3) the proportion of NMB attributable to United Kingdom research; and (4) the elapsed time between research funding and health gain. The data collected from these four key elements were used to estimate the internal rate of return (IRR) from MSD-related research investments on health benefits. We analysed the uncertainties in the IRR estimate using a one-way sensitivity analysis.

Results: Expressed in 2013 prices, total expenditure on MSD-related research from 1970 to 2013 was £3.5 billion, and for the period used to estimate the rate of return, 1978-1997, was £1.4 billion. Over the period 1994–2013 the key interventions analysed produced 871,000 QALYs with a NMB of £16 billion, allowing for the net NHS costs resulting from them and valuing a QALY at £25,000. The proportion of benefit attributable to United Kingdom research was 30% and the elapsed time between funding and impact of MSD treatments was 16 years. Our best estimate of the IRR from MSD-related research was 7%, which is similar to the 9% for CVD and 10% for cancer research.

Conclusions: Our estimate of the IRR from the net health gain to public and charitable funding of MSD-related research in the United Kingdom is substantial, and justifies the research investments made between 1978 and 1997. We also demonstrated the applicability of the approach previously used in assessing the returns from cardiovascular and cancer research. Inevitably, with a study of this kind, there are a number of important assumptions and caveats that we highlight, and these can inform future research.

Keywords: Medical research investment, QALYs, Musculoskeletal disease, Medical research charities, Value of health, Rate of return, Elapsed time, Research payback

* Correspondence: jonathan.grant@kcl.ac.uk
[2]Policy Institute at King's, King's College London, Virginia Woolf Building, 22 Kingsway, London WC2B 6LE, United Kingdom
Full list of author information is available at the end of the article

Background

Total global investment in biomedical and health research was estimated at US$240 billion in 2009 [1], equivalent to approximately US$270 billion in 2016. These investments are intended to improve health for patients and the public. But do they? And if so, what are their returns?

In recent years, researchers and research funders have aimed to better understand the range of impacts arising from public and charitable funding for medical research, including the resulting economic benefits. Such information provides accountability to taxpayers and charity donors, and increases our understanding of how research effectively translates to health gains. In this paper, we examine the economic returns from musculoskeletal disease (MSD) research. This is the third in a series of studies looking at the returns from cardiovascular (CVD) research [2] and cancer research [3], as well as the broader economic impacts or spillover effects of research funding [4].

As reviewed by Buxton et al. [5], and updated by Glover et al. [3] and Raftery et al. [6], the literature that assesses the value of the benefits of medical research forms a relatively limited field in terms of methodology and quality. There are two broad approaches. Firstly, a 'top down' approach where overall health gains in a disease area are related to research investments, but this requires an estimate of how much of the total health gain can be attributed to medical research investments. For example, Funding First [7] argued, in a report entitled 'Exceptional Returns', that the steep decline in CVD deaths in the United States between 1970 and 1990 had an economic value of US$1.5 trillion annually, and deduced that one-third of this (US$500 billion a year) could be attributed to medical research that led to new procedures and drugs. The approach was replicated in a series of studies by Access Economics [8, 9] and Deloitte Access Economics [10] estimating the return on Australian biomedical research on the basis of overall improvements in Australian lifespan. The base-case assumption in these studies was that research was responsible for 50% of the improvements in healthy lifespan, although it is worth noting that the authors acknowledged there was no evidence to support this assumed rate of attribution.

The challenge of top-down attribution can be addressed by examining in a 'bottom-up' manner the impacts of specific projects or programmes of research by tracing forwards from the research to the benefits that arise. This is the approach developed by HERG [2] and Glover et al. [3], and adopted in this study. Here, we estimate the net monetary benefits (NMB), defined as the health benefit valued in monetary terms minus the cost of delivering that health benefit, for a set of key interventions to reduce MSD that arose from the United Kingdom application of relevant United Kingdom research. This 'bottom-up' approach led to an impressive but less 'exceptional' internal rate of return (IRR) of 9% and 10% for CVD and cancer research, respectively [2, 3].

However, in both these studies the reduction of smoking over the period analysed had a major impact on the estimated rate of return. For example, the return on cancer research investment declined to 2.4% in a sensitivity analysis that excluded the effect of smoking cessation, and attribution of the reduction in smoking to medical research alone is contestable. Glover et al. [3] therefore concluded it would be valuable to undertake an investigation in another clinical area, such as MSD, in which smoking only marginally affects outcome to see whether similar rates of return are found [3] (smoking has a comparatively small effect on the musculoskeletal system, including a reduction in bone mineral content and deleterious effects on osteoporosis, fractures and other MSD [11, 12]). Another prima facie, methodological reason why MSD research is an interesting case to examine is that it largely relates to chronic conditions, where health gains occur through improvements in morbidity, rather than mortality as was the case for CVD and cancer.

The MSD burden of disease and research

How much biomedical and health research funding is invested in different disease areas is determined by a number of factors, including burden of disease, scientific tractability, donor appeal and previous investment [13]. The United Kingdom Clinical Research Collaboration [14] report a relatively weak correlation between research investment and burden of disease (using disability adjusted life years (DALYs)), using the health categories in the Health Research Classification System (HRCS).[1] Whilst 'cancer' has the highest proportion of spend and highest DALY rate (ca. 20%), the combined health research categories 'cardiovascular', 'blood' and 'stroke', have approximately 16% of burden, but only 9% of the spend. 'Musculoskeletal' has an even greater skew, with approximately 9% of the burden but only 3% of spend.

MSD has relatively low rates of mortality, although evidence indicates incidences of deaths in which MSD conditions were the underlying cause of death are under-reported [15, 16]; however, it has a relatively high prevalence of disability and morbidity. Many musculoskeletal conditions are recurrent and lifelong disorders which can often cause long-term pain, physical disability, loss of independence, reduced social interaction and a decline in quality of life [17]. Arthritis conditions, for example, are the biggest cause of pain and disability in the United Kingdom [16].

While most MSD conditions do not require hospital admission, MSDs are a frequent cause of consultation

with general practitioners (GP). For example, 15–20% of all GP consultations involve a patient with MSD conditions [17]. Further, Woolf et al. [18] found, in a cross-national comparison, that MSD conditions are one of the leading causes of both long-term absences from work and disability pension claims.

MSD conditions affecting joints, bones, muscles and soft tissues can affect any age group, but the prevalence of the disease increases drastically for older people. The age group most commonly affected (50+ years old) tends to fall predominately outside of the active labour force. Further, conditions which fall under MSD affect approximately 10 million people in the United Kingdom, accounting for £5 billion of the NHS programme budget spend in England alone [16].

Therefore, MSD has a very different funding and disease profile to that of the two previous studies on CVD [2] and cancer [3].

Defining the scope of MSD

For this study, we needed a clear and internationally defensible definition of 'musculoskeletal' disease. Following consultation with a number of experts (see acknowledgements) we used Chapter XIII of the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), known colloquially as ICD 10 Chapter XIII [19]. One advantage of using Chapter XIII of ICD 10 is that it is also the basis of the musculoskeletal category in the HRCS, meaning that, in many cases, research investment and health outcomes are defined using the same criteria.

As discussed in more detail below, we focussed on five condition groups/areas (with the number indicating the ICD sub-classification):

- Inflammatory arthritis (M00–M14): particularly rheumatoid arthritis (RA), juvenile idiopathic arthritis (JIA), ankylosing spondylitis, psoriatic arthritis and gout
- Osteoarthritis (M15–M19)
- Connective tissue disorders (M30–M36): particularly systemic lupus erythematosus (SLE) and dermatomyositis
- Back pain and dorsopathies (M40–54)
- Osteoporosis (M80–82)

Methods

Overall approach

The overall conceptual approach is summarised in Fig. 1 and requires four key data elements to estimate the IRR arising from MSD research, namely (1) a time series of public and charitable funding of MSD-related research; (2) a time series of NMB of MSD health gains, derived from the monetised health benefits and healthcare costs from the actual use of selected interventions; (3) an estimate of the elapsed time between the investment (research funding) and return (health gain) associated with those interventions; and (4) an estimate of the amount of health gain that should be attributed to United Kingdom public and charitable research investment in MSD-related research.

With these four data inputs, the IRR on the public and charitable investment in MSD research and development (R&D) can be calculated (it should be noted that the costs of private sector R&D investments are accounted for in our analysis as elements within the cost of delivering healthcare, which are netted off in the NMB). The costs to the health service of medical interventions produced by the private sector are assumed to include the return to the private sector on its R&D investments.

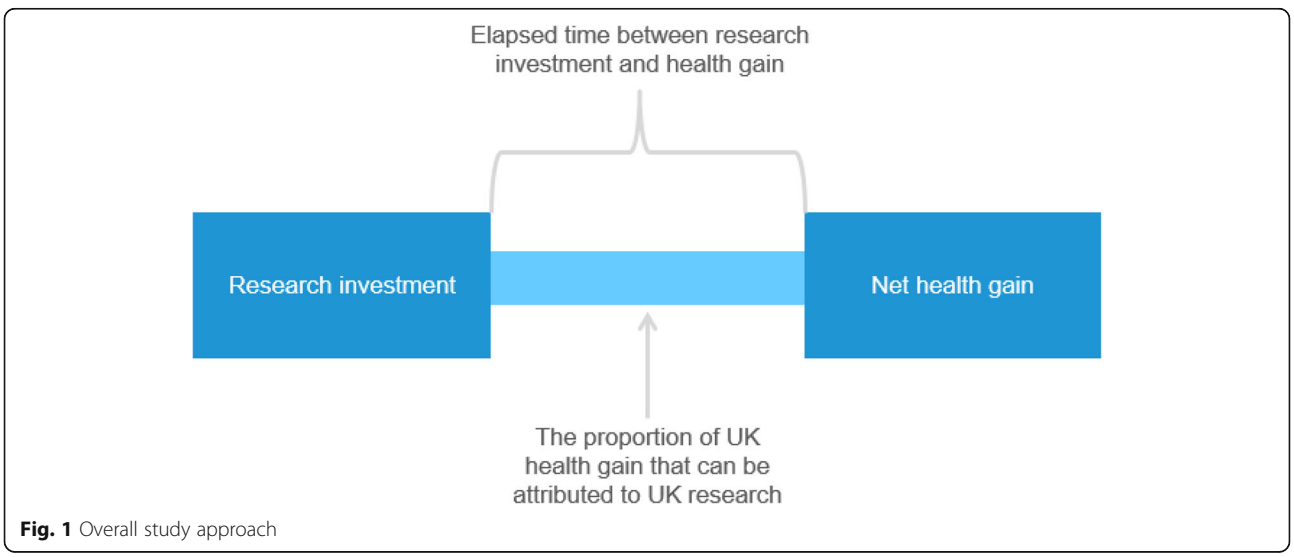


Fig. 1 Overall study approach

Estimating public and charitable funding of MSD-related research

We developed a time series of public and charitable funding of MSD-related research between 1970 and 2013 for the five largest research funders and a group of other research charities. Each involved a different approach.

Medical Research Council (MRC)

The MRC had previously provided digital copies of its annual reports dating back to 1911. Between 1976 and 1992, the MRC used a consistent disease classification system for its research grants, including the category 'Muscle, Bone and Joints', which we used for this study. Data from 1976 to 1992 was extracted from heading 2 of the annual reports, which classifies projects according to relevance – that is, the total spend on each project is placed against any (and all) relevant categories (i.e. it can be double counted). In comparison, this classification method is more 'inclusive' than the alternative first heading in the MRC annual reports, which uses a classification according to the primary purpose, that is, the total spend equates to the total MRC spend for the year. As with the previous studies we use this broader definition of expenditure as it is likely to overstate funding and thus err on the side of being conservative when calculating the rate of return. In addition, the MRC gave access to annual reports detailing programme expenditure on conditional area 'Musculoskeletal' research from 2009 to present day, based on the HRCS definition.

Information was not available between the periods of 1970 to 1975 and 1993 to 2008, in which case we calculated missing data by inter- and extrapolating the missing data using various growth functions in Excel.

Wellcome Trust (WT)

The WT produced a detailed list of the grants awarded from 1970 to present day, and a summary of the total commitment annually. The WT grant management system identified awards with a 25% or higher proportion classified as 'musculoskeletal', based on the HRCS definition.

Once data was compiled in Excel it was analysed for patterns and irregularities. For example, in 2008, there was an increase of approximately £28 million in total commitment as a result of grants awarded for three long-term research programmes, all in biomedical engineering. Upon request, WT provided further grant descriptions on the proportion of grants which had been classified as less than 50% 'Musculoskeletal' to determine whether the grant was within scope of the disease area as defined by ICD 10, and we reduced the total commitment to reflect the proportion of the grant falling under this classification.

To adjust the total commitment data into total expenditure, we assumed an average of 3 years for each grant and allocated commitment over this time period and re-calculated expenditure on a per year basis.

Arthritis Research United Kingdom (ARUK)

We assumed all ARUK research expenditure related to MSD. ARUK provided us with detailed data on the total commitment as recorded in their income and expenditure statements, up to and including 2008, and total expenditure from 2009 onwards.

Once data was collected and collated into an Excel spreadsheet, it was analysed for irregularities. The data showed a significant decrease in 2002, which mainly resulted from a change in research strategy, and therefore a pause in the funding of new grants.

As with the Wellcome Trust we adjusted commitment data (between 1970 and 2008) by assuming an average grant length of 3 years and allocating expenditure per year on that basis.

MSD research activity index

The Department of Health (DH) and Funding Councils (FCs) did not record information on research funding by disease area. As with the previous studies, we were able to generate a total expenditure for both, as described below, and multiplied this by an 'activity index' to estimate the amount of research expenditure on MSD annually.

The activity index was estimated by looking at the total expenditure on MSD research by the MRC and the WT and by comparing it with bibliometric data that was commissioned to inform other elements of the study. We also compared it to other sources, including a historical analysis of NHS research, which suggests that 4.5% of research outputs were related to MSD research [20], and a more recent analysis using the HRCS, which suggests that 2.8% of research spend by the top 12 public funders in the United Kingdom is on MSD research [14]. Overall, we assumed that 3% of all biomedical and health research activity is related to MSD research, ranging between 2% and 4% for sensitivity analysis.

Department of Health (DH)

The DH did not have information available on total MSD research spend, nor did they have data on the total spend from one source. Additionally, we were interested in estimating the total research spend by the DH as well as the National Health Service (NHS), collectively. Therefore, as with the previous studies, we collected data from three sources, namely data for 1973 was entered by hand from Maddock, 1975 [21]; data for 1981 to 1984 was entered by hand from the Annual Review of Government Funded R&D, 1984; and data on the DH (excluding NHS) from 1986 onwards data were collected from SET statistical table 3.1 for Department

of Health and Social Security and the DH. NHS funding was included from 1995 onwards and subsequent NIHR data from its founding in 2006.

Data was not available for either NHS or the DH for 1970–1972 or 1974–1980. Further, data for the NHS could not be extracted prior to 1995. Therefore, we estimated the expected funding in Excel for both NHS and DH separately to provide a time series for each, and added the two estimates for the total expenditure. We then multiplied our total DH/NHS/NIHR funding series by the activity index (as described above) to generate an estimate of total DH funding in MSD research.

Funding Councils (FCs)

Similar to the DH, the FCs did not differentiate research spend by field area, and therefore data on total research spend was collected and/or estimated and multiplied by the research activity index (see above). Data was collected from 1989 onwards from three sources, namely for 1989–1992 from the Research Grant figures for Great Britain, provided by the HEFCE[2]; funding from 1993 to 2008 was extracted from the HEFCE mainstream quality-related research grant allocations for biomedical subjects in the years 1993–1994 to 2008–2009[3]; and funding from 2009 to 2012 was provided through the HEFCE mainstream quality-related research grant allocations for biomedical subjects in the years 2009–2010 to 2012–2013.[4]

Data could not be extracted from 1970 to 1988; therefore, we took a similar approach to that used for DH funding. We projected the best linear fit of data for the period 1988 to 2012, then determined the expected growth from the same time series in order to estimate for missing annual data.

Other medical research charities

In addition to WT and ARUK, we were aware of other medical research charities that supported MSD research. We therefore approached the Association of Medical Research Charities (AMRC), which helped us identify and select that 'other group'. Using the HRCS report for 2014 [14], AMRC identified 21 members who funded MSD research. Two of these charities were out of scope (because they were funding research outside the United Kingdom), leaving 19 AMRC members with a total spend of approximately £18 m per year on MSD research. When WT and ARUK were excluded, the remaining 17 charities spent approximately £2 m annually, with the top nine funders of the remaining charities accounting for 96% of this investment. We therefore asked these nine other charities for funding data back to 1970. In many cases, the charities did not have sufficiently robust data management systems to go back that far, and in a number of cases were established at some point during our time series. Furthermore, some had different financial years and different accounting practices (i.e. commitment of multiple year research funding vs. in year expenditure). One charity declined to participate on the grounds that it did not have the resources to collate the information. We worked closely with the other charities to develop our best estimated time series and combined this as 'other medical research charities' in our analysis. We deliberately present the aggregate data to protect the confidentiality of the charities and the data they provided. Overall the 'other medical research charities' account for approximately 4% of total expenditure on MSD research. For the sensitivity analysis, and to take into account missing data, we increased the other expenditure for the 'other medical research charities' by 20% for our high estimate.

Taking inflation into account

To calculate the total cumulative spend in real terms, the total nominal research spend was adjusted for inflation. We applied a Gross Domestic Product deflator sourced from the HM Treasury (base year = 2013/14) and adjusted total spend for each year based on this [22]. Thus, cumulative funding over the period we examined is expressed in 2013–2014 GBP.

Royalty payments

In principle, any royalty payments received by research funders as a result of their research investment in the relevant time period should be netted off in the year they occur and so reduce the present value of the investment stream. In previous studies, we had no evidence to suggest that such royalty payments would be sufficient to make a substantive difference to the estimated rate of return. In this case, returns from the royalties relating to the commercial development of anti-tumour necrosis factor (anti-TNF) drugs were believed to be sufficient to have an impact on the IRR. We accessed data from the published annual accounts of the Kennedy Trust and data supplied by ARUK to illustrate, in a sensitivity analysis, the magnitude of the effect of these royalties.

Estimating the NMB from MSD-related research

This element of the study required estimates of the lifetime quality adjusted life years (QALYs) gained and the net lifetime costs to the NHS of delivering those QALYs for relevant research-based interventions provided in each year of the period 1994–2013. Incremental QALYs encompass both survival and quality of life gains from an intervention as compared to prior practice. We used QALYs gained to quantify health gain rather than changes in DALYs. Although DALYs are used in much of the literature on overall burden of disease itself to characterise population health loss, QALYs are the more

appropriate (and much more commonly used measure, particularly in the United Kingdom) to characterise the gain from the use of specific interventions. As far as data permitted, the methods and sources used were chosen to provide directly comparable results to those in the two previous studies on the returns on investment on CVD and cancer research [2, 3].

Overall estimates of the QALYs, and the costs to the NHS of delivering them, were built-up by aggregating estimates for a series of specific interventions. As before, this approach required identification of the key relevant MSD interventions and the number of new patients actually receiving them in the NHS in each year of the relevant period and estimates of the discounted life-time QALY gains and net life-time costs per patient resulting from initiation of the intervention. The aggregated QALYs gained were then valued in monetary terms using, as before, a base-case opportunity cost value of a QALY to the NHS of £25,000 – the midpoint value of the National Institute for Health and Care Excellence's (NICE) threshold range [23]. From this, the similarly aggregated discounted net lifetime NHS costs of delivering that health benefit were deducted to provide the overall estimate of the NMB. Any specific circumstances where data limitations forced deviation from this approach are noted below.

In the absence of any study that had identified and quantified the research-based MSD interventions that had, during the relevant period, contributed most to the United Kingdom population health gain in this area, or to substantial changes in costs, we reviewed sources that might help build an initial view of likely interventions that might be included. Particularly important in this stage were studies identifying the burden of MSD disease in the United Kingdom [24], relevant NICE Pathways [25], and NICE and National Collaborative Centre (NCC) Guidelines [26–29]. With the assistance of ARUK, we then identified key experts (see Acknowledgements) who, through a workshop (November 2015) and subsequent direct one-to-one interactions, helped produce a list of interventions that, in principle, looked appropriate for inclusion. More detailed review of available data and cost-effectiveness evidence was undertaken to confirm the importance of the listed interventions and to establish whether the necessary estimates of net costs and benefits, and levels of usage, were available. Further input from experts was sought (November 2016 to January 2017) to confirm our assumptions, check for any perceived omissions and to validate the emerging findings, after which some final adjustments were made.

QALY gains and costs of chosen interventions

We identified appropriate published studies that had estimated the cost-effectiveness of the chosen interventions in the United Kingdom. Wherever possible, we used independent studies produced for NICE or for national clinical guidelines and published in the Health Technology Assessment monograph series, or estimates that had been reviewed and accepted by NICE. In some instances, evidence was taken directly from NICE Technology Appraisals or from National Collaborating Centre Guideline modelling. Where more than one relevant study had been undertaken for NICE, we used the most recent to reflect the developing evidence base. Where no such study for NICE was available we sought the most relevant United Kingdom focussed study from published literature.

Constructing a time series (1994 to 2013) of usage of MSD interventions

To estimate total NMB for the period, per-patient QALY gains and net costs for each intervention were multiplied by the estimated number of new patients who actually received each intervention in each year. We used the following methods to estimate the time series of usage for the selected interventions.

Data on the number of patients receiving procedural interventions (e.g. hip replacements and surgical length of stay) were gathered from Hospital Episodes Statistics [30] available for years 1999–2013. For pharmacological interventions, prescribing data on total annual spend in the NHS over the period of interest was utilised. Two primary sources of net ingredient cost (NIC) were available – (1) Prescription Cost Analysis (PCA) [31] and (2) Hospital Prescribing England (HPE) [32]. Information on the average cost of a regimen was used, as well as accounting for usage across different diseases and indications, to estimate the number of patients receiving the intervention. Finally, estimates of the average duration of treatment allowed an estimation of the number of patients starting treatment in any of the given years of interest. These data were publically available over different time periods (PCA 1998–2013; HPE 2004–2013). Where an intervention was launched before available data, but within the time period, a linear interpolation was performed with usage assumed to be zero the year before launch. Where an intervention was launched prior to 1994 the last known value was carried back. For years where only PCA data were available, but prescribing also occurred in secondary care, a ratio of the last year of available PCA and HPE data was used to uprate years with PCA NIC only. Estimating the number of patients receiving interventions for low back pain (LBP) was approached differently, using general practice data provided for this study from the Clinical Practice Research Datalink (CPRD) [33].

The component figures of numbers of people receiving treatment interventions were mainly derived from data for England. To produce a United Kingdom estimate

(needed because research spend data is for the United Kingdom) figures were adjusted by a factor reflecting England's proportion of the adult United Kingdom population [34]. All cost estimates were adjusted to 2013–2014 prices using the Hospital and Community Health Services Pay and Prices Index [35].

The value placed on the estimated QALYs gained is a critical parameter in estimating the return on research investment. Given that public spending on health research (whether from taxation or from public donations to medical charities) can be seen as a decision to achieve health benefits through research rather than directly through current healthcare, a value for a QALY (resulting from research investment) should arguably reflect the marginal opportunity cost of generating QALYs in the NHS. In the calculation of NMB in this study, as in the previous two studies, we used as the base-case value an operational opportunity cost value of a QALY in the middle of the 'threshold range', as used by NICE in its Technology Appraisals of £25,000 [23]. However, this value can be contested; on the one hand, detailed econometric analysis has estimated that the marginal opportunity cost value in recent years has been significantly lower, at approximately £13,000 [36]. On the other, the value that society places on a QALY as recommended for use in quantifying the impacts of government policies is estimated to be £60,000 [37]. In addition to the base-case, we report values from £13,000 to £60,000.

Analysis of United Kingdom clinical guidelines to estimate elapsed time and rate of attribution

In the previous studies on CVD [2] and cancer [3] research, the references cited in a sample of clinical guidelines were analysed to inform the estimate of the elapsed time between research spend and net health gain, and the proportion of net health gain that could be attributed to United Kingdom research. In the current study on MSD research, we replicated this approach.

In line with the process for identifying musculoskeletal interventions, guidelines were identified and classified in terms of their relevance for inclusion by comparison to Chapter XIII of the ICD-10 disease classification [19]. Based on this inclusion criterion, a total of 22 national guidelines were identified, spanning a range of practice in the field and issued by ten different bodies (Table 1).

We used a bespoke computer programme to extract references from the electronic PDF version of each guideline. In seven cases, the automated reference extraction failed (because papers were not referenced in a recognised format). In these instances, references were extracted manually.

Of the 3640 references cited in the 22 national guidelines, 2746 references (75%) were extracted automatically and 894 (25%) manually. References from non-journal sources (which were unlikely to constitute original research) and duplicates within the same guideline were removed, leaving a total of 3237 references. The average age of the papers cited in a clinical guideline has been termed the 'knowledge cycle time' [38], which is the average difference between the publication date of the clinical guideline and the publication date of papers cited in the guideline. The knowledge cycle time was calculated for the 22 identified national guidelines, and used to inform the estimated elapsed time.

To estimate the rate of attribution to the United Kingdom, the 3640 extracted references were provided to the Centre for Science and Technology Studies (CWTS)[5] to be matched to their bibliometric database (which is derived from the Web of Science). Of the extracted references, CWTS was able to match 2804 (84%); 40 additional references were manually matched, for a total of 2844 (85%). Address data was successfully retrieved from Web of Science for 2762 of these papers. This dataset was used to estimate the degree of attribution to the United Kingdom, based on the addresses of all authors of the included papers. These addresses were used as a proxy for the location in which the research was conducted, and so it was possible to estimate the proportion of the research cited in guidelines that was conducted in the United Kingdom. The non-matched references included non-serial outputs, such as books and websites, journals that are not indexed in the Web of Science, papers whose publication pre-dates a journal's indexation in Web of Science and incorrect references.

Calculation of the rate of return

Using the four key sources of data summarised in Fig. 1, we can attribute a proportion of the estimated total annual NMB of the MSD health gain as being due to United Kingdom research, and relate an equal number of years of investment to years of NMB, 'lagged' by an estimate of the elapsed time between research and benefit. As in the previous studies, we express this return on investment as an IRR, which is effectively the discount rate that would yield a zero net present value. In this application, the formula for the IRR is:

$$-\sum_{t=1}^{20}\frac{ResInv_t}{(1+IRR)^t}+\sum_{t=1+Lag}^{20}\frac{NMB_t(Attrib)}{(1+IRR)^t}=0$$

Where, *Res Inv* is the United Kingdom research spend on MSD in year *t*, *NMB* is the net monetary benefit in year *t* (monetary value of QALYs gained minus costs of delivery), *Lag* is the estimated average years between research spend and health gain, *Attrib* is the proportion of

Table 1 Summary of United Kingdom guidelines included in analysis of elapsed time and attribution

Provider	Guideline	Year
British Association/College of Occupational Therapists	Hand and wrist orthoses for adults with rheumatologic conditions: practice guideline for occupational therapists	2015
British Association/College of Occupational Therapists	Occupational therapy for adults undergoing total hip replacement: practice guideline for occupational therapists	2012
British Pain Society	Guidelines for pain management programmes for adults	2013
British Pain Society	The assessment of pain in older people	2007
British Society for Rheumatology	British Society for Rheumatology and IASP Musculoskeletal Pain Taskforce Guidelines for the integrated management of musculoskeletal pain symptoms	2008
British Society for Rheumatology	British Society for Rheumatology guidelines on standards of care for persons with rheumatoid arthritis	2005
National Institute for Health and Care Excellence	Osteoarthritis (CG.177)	2014
National Institute for Health and Care Excellence	Osteoporosis (CG.146)	2012
National Institute for Health and Care Excellence	Hip fracture (CG.124)	2011
National Institute for Health and Care Excellence	Rheumatoid arthritis in adults: management (CG.79)	2009
National Institute for Health and Care Excellence	Low back pain in adults (CG.88)	2009
National Osteoporosis Foundation	Clinician's guide to the prevention and treatment of osteoporosis	2014
National Osteoporosis Guideline Group	Guideline for the diagnosis and management of osteoporosis in postmenopausal women and men from the age of 50 years in the United Kingdom	2014
National Osteoporosis Guideline Group	Osteoporosis: clinical guideline for prevention and treatment: executive summary	2014
National Osteoporosis Society	Vitamin D and bone health: a practical clinical guideline for patient management	2013
Royal College of Nursing	Administering subcutaneous methotrexate for inflammatory arthritis	2013
Royal College of Physicians	Pain: complex regional pain syndrome	2012
Royal College of Physicians	Upper limb disorders: occupational aspects of management 2009	2009
Scottish Intercollegiate Guidelines Network	Management of osteoporosis and the prevention of fragility fractures (CG.142)	2015
Scottish Intercollegiate Guidelines Network	Management of chronic pain (CG.136)	2013
Scottish Intercollegiate Guidelines Network	Management of early rheumatoid arthritis (CG.123)	2011
Scottish Intercollegiate Guidelines Network	Management of hip fracture in older people (CG.111)	2009

NMB attributed to United Kingdom research and *IRR* is the internal rate of return.

The IRR is convenient in enabling a comparison to be made between non-competing investments of different sizes with different start dates, as well as providing a direct comparison with our previous work.

Given the nature of the numerous necessary judgements involved, the multiple sources of evidence, the multiple parameters, and the many and various layers of estimates and assumptions, a probabilistic sensitivity analysis was not conducted. Even if it were possible, it would not be informative to express all the uncertainty as ranges for each parameter to reflect stochastic uncertainty in our overall estimate. Instead, we present a range of one-way sensitivity analyses that provide an indication of the uncertainty associated with each of the key aggregate parameters that goes into the calculation, namely the size of the research investment, the average elapsed time between research spend and use of the intervention, the magnitude of the NMB, the proportion of the NMB that can be attributed to United Kingdom research, and the effect of netting-off royalty payments from the investment stream.

Results

Public and charitable funding of United Kingdom MSD-related research, 1970–2013

Additional file 1 provides a detailed account of the estimated total expenditure by year by organisation over a

43-year period (1970–2013). A summary of cash expenditure by funder can be found in Fig. 2. The significant spike in 2008 can be attributed to three significantly large grants committed by the WT to the development of large facilities and long-term research programmes. Figure 3 shows the estimated public and charitable expenditure on MSD-related research as £1.4 billion from 1978 to 1997 in constant 2013 prices (i.e. adjusted for inflation). As noted below, 1978–1997 is the funding period used to calculate the IRR taking into account the estimated elapsed time. Figure 3 presents a sensitivity analysis with 'high' and 'low' scenarios for total MSD-related research spend, with a range of £1.2 billion to £1.6 billion.

We additionally obtained two sets of estimates of royalty payments arising from anti-TNF commercialisation since 2002. The first set was for total royalty payments to the Kennedy Trust and the second was for the sum of the royalties retained by the Trust and those remitted to ARUK. The first is likely to be an over-estimate for our purposes as it includes some royalties received by private individuals; the second may underestimate the total sum that returned into medical research spending. We used both in the sensitivity analyses.

Interventions

A broad review of the field and discussions with our Advisory Board led us to focus on five main disease areas in which it appeared that the most significant research-based changes to healthcare delivery had occurred between 1994 and 2013. These were inflammatory arthritis, osteoarthritis, connective tissue disorders, osteoporosis and back pain.

Inflammatory arthritis (M00–M09: RA, JIA and psoriatic arthritis; M10–M12: Gout; M45: Ankylosing spondylitis)

Key interventions:

- Early, aggressive, combination therapy
- Use of anti-TNFs (infliximab, etanercept, adalimumab, golimumab, certolizumab)
- Use of other biologics (tocilizumab, abatacept, rituximab)
- Allopurinol and febuxostat in treatment of gout

The management of RA and other associated inflammatory arthritis has evolved over the studied time period, predicated by the advent of disease-modifying anti-rheumatic drugs (DMARDs). Conventional DMARDs (cDMARDs), most notably methotrexate, are now a standard component of initial RA management. Treatment has shifted from monotherapy or slow step-up regimen, towards early and aggressive combination therapy (EACT).

Use of methotrexate served as a proxy for all conventional DMARD therapy adjusting for a proportion of methotrexate co-prescribing with biologic DMARDs (bDMARDs). Methotrexate NIC data was only available from the PCA, so the estimates of usage may constitute a slight underestimate. However, given that maintenance

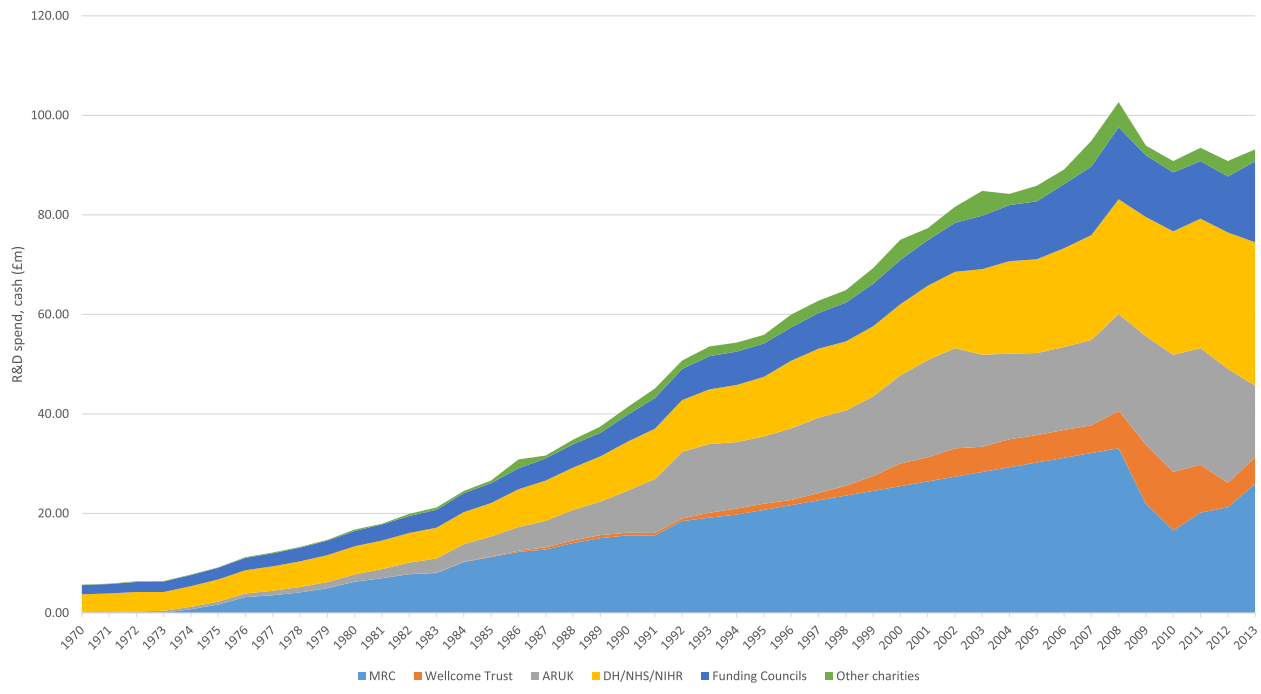


Fig. 2 Cash expenditure on MSD research from 1970 to 2013, by funder

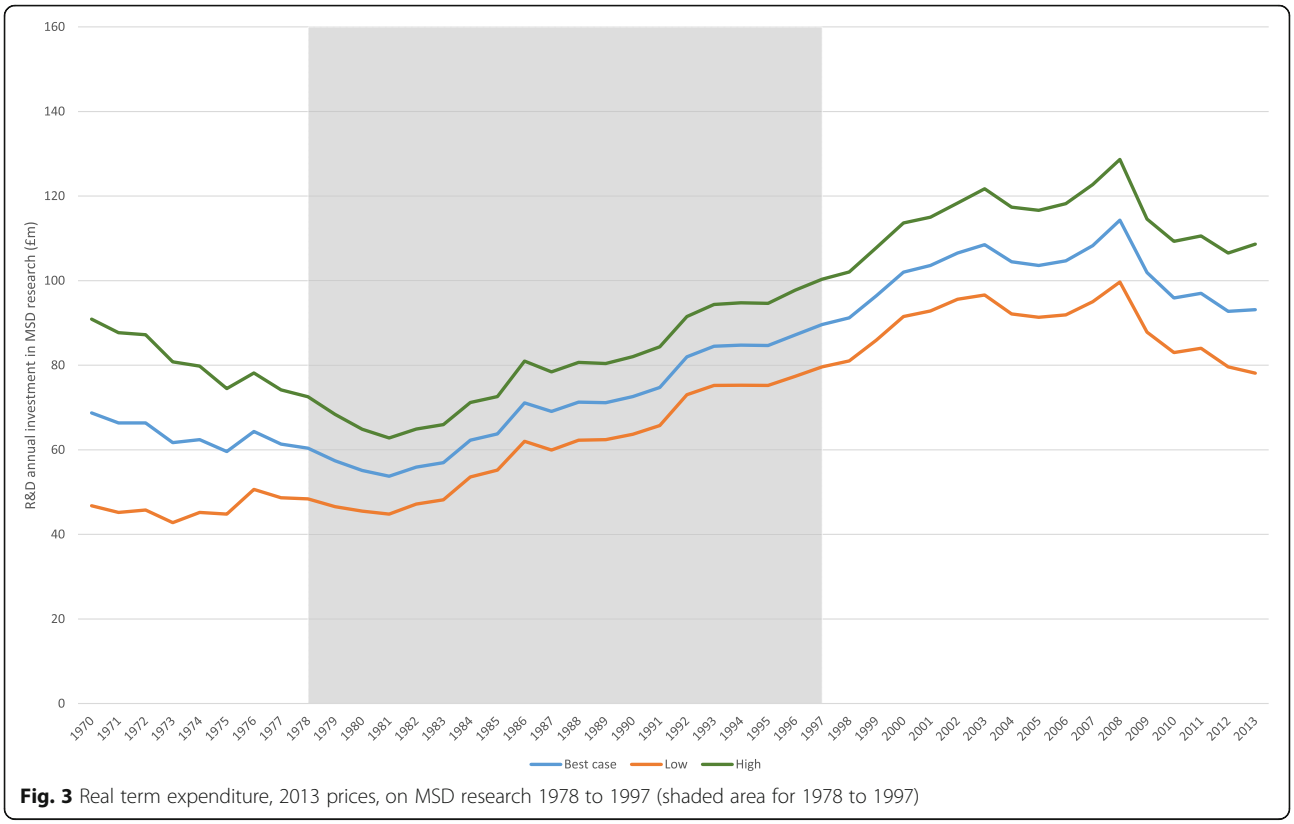


Fig. 3 Real term expenditure, 2013 prices, on MSD research 1978 to 1997 (shaded area for 1978 to 1997)

doses are normally prescribed in primary care and that patients are likely to receive treatment for 10 years on average [39], most prescribing will be captured in primary care prescribing data. Clinical experts estimated that 40% of patients were managed with EACT in 1994 and essentially all by 2000.

Since the early 2000s the major new treatment option of bDMARDs became available for patients with more severe RA. Anti-TNFs were the initial generation, with other new bDAMRDs emerging that affect RA through other mechanisms.

Estimates for the use of both cDMARD and bDMARD therapy took account of their use in other disease areas (ulcerative colitis, Crohn's disease, psoriasis, SLE) and across inflammatory arthritis indications (where net health gains are likely to differ). Usage was split proportional to incidence. NICE estimates were used to further allocate bDMARD use across different treatment stages (bDMARD naïve, after failure of an anti-TNF).

Cost effectiveness evidence was available from Tosh et al. [40] for EACT, and was inferred from Stevenson et al. [41] for bDMARDs for RA patients not previously treated with biologics or having failed on cDMARD therapy. Studies by Malkotti et al. [42], Jackson et al. [43] and Minton et al. [44] provided estimates of bDMARD net health gains for patients previously treated with DMARDs including an anti-TNF.

In the absence of indication specific cost-effectiveness data for EACT, its benefits were assumed to be the same for the JIA population and psoriatic arthritis. Shepherd et al. [45] provided estimates of bDMARD net health gains after the failure of cDMARD management and after the failure of an anti-TNF for a JIA population. Rodgers et al. [46] and Corbett et al. [47] were used for bDMARD net health gains in psoriatic arthritis and ankylosing spondylitis, respectively. Beard et al. [48] provided cost-effectiveness data on allopurinol and febuxostat for the treatment of gout.

Osteoarthritis (M15–M19)

Key interventions:

- Move to cementless and hybrid hip prostheses
- Use of minimally invasive hip and knee replacement
- Decreased hospital length of stay for hip and knee replacement from change in surgical management and early rehabilitation
- Use of Cox-II inhibitors (celcoxib, etoricoxib, meloxicam, etodolac)
- Concomitant Cox-II inhibitor use of proton pump inhibitors (lansoprazole, omeprazole, esomeprazole pantoprazole)

Joint replacement has constituted a mainstay of treatment to alleviate pain and regain function of damaged joints caused by osteoarthritis for some time. As such, it was not appropriate to include all benefits

from hip and knee replacement during the period of interest. Incremental changes associated with the use of minimally invasive techniques were, however, relevant to this period, as well as a trend towards cementless and hybrid prostheses for hip replacement. There has also been a marked reduction in hospital length of stay for patients undergoing joint replacement. Mean hospital length of stay for hip and knee replacement surgery in 1999 was 12.7 and 12.3 days, respectively. By 2013, these figures were 5.7 (hip) and 5.1 (knee).

Data from the National Joint Registry [49] was used alongside Hospital Episodes Statistics to estimate the number of minimally invasive hip and knee replacements and type of prostheses used in hip replacement.

Net health gains for cementless and hybrid hip replacements were taken from Pennington et al. [50] and minimally invasive joint replacement from de Verteuil et al. [51]. All hip replacement net health gains were attributed to osteoarthritis, but include a very small proportion of replacement due to other reasons such as RA or dysplasia of the hip. DH reference costs were used to assign a unit cost to a 1-day reduction in length of stay. Savings were estimated by multiplying the annual number of replacements by the difference in length of stay compared to baseline.

All Cox-II inhibitor use was assumed to be in osteoarthritis, with net health gains attributed as such, using evidence from NCC modelling [52]. Cox-II inhibitors taken off the market (i.e. Vioxx) were overall assumed to produce no net benefit. A proportion of concomitant proton pump inhibitor use was assumed over the period, starting at 0% in 2006, rising to 30% by 2012 [52].

Connective tissue disorders (M30–M35)

Key interventions:

- Mycophenolate mofetil for SLE

Net health gains resulting from mycophenolate mofetil were available for treatment of SLE [53], although limited to a nephritis population.

Osteoporosis (M80–M82)

Key interventions:

- Bisphosphonates (alendronate, etidronate, risedronate, zoledronate)
- Hormonal therapies and dual action bone agents (raloxifene, teriparatide, denosmuab, strontium ranelate)

Bisphosphonates, the first of which was launched in the mid-1990s, are recommended as a first-line treatment for post-menopausal osteoporosis and can be used in both primary and secondary fracture prevention. Subsequent hormonal therapies have been developed, which have similar properties.

Some therapies have become generic and these price changes were reflected in the average cost of a regimen over time to estimate the number of patients receiving treatment. However, such changes are not reflected in the cost-effectiveness evidence, and thus will tend to overestimate the lifetime costs of delivering healthcare. Data on zoledronate (Aclasta) NIC was not available in HPE and the manufacturer provided some internal data.

Stevenson et al. [54] provided estimates of per patient net health gains for most of these treatments, although evidence from NICE TA204 2010 [55] was used for zoledronate and denosumab. These data were provided split by age group, and as such net health gains were weighted to reflect the age distribution of the United Kingdom population. Assumptions about the proportion of osteoporosis intervention that is aimed at primary and secondary fracture prevention were taken from NICE estimates [55, 56].

Back pain (M54)

Key interventions:

- Manual therapy
- Structured exercise programmes
- Combined psychological and physical therapy

The focus was primarily on chronic low back pain and sciatica as defined by NICE guidance [57]. In the absence of a comprehensive source through which to ascertain the number of physical and psychological interventions for LBP that patients have received over the period of interest, we had to use a different approach to estimating the population.

Data on new diagnoses of LBP were obtained from the CPRD, which provides observational data from United Kingdom GP practices. Of the Read codes used to identify relevant LBP, approximately 95% of events were one of the following: LBP, back pain without radiation not otherwise specified, sciatica, complaint of LBP, pain in lumbar spine, mechanical low back pain, chronic low back pain, back pain, unspecified, or lumbago. Incidence was defined as the number of incident events divided by the total registered CPRD population (person years) after removing participants who had ever had previous back pain as well as the first year of CPRD sample follow-up, who were defined as not 'at risk'.

Incidence was split into sex-specific 5-year age bands. United Kingdom population figures were used to estimate a total number of incident cases of back pain during the period based on the CPRD sample. The focus was on a chronic population who receive active

interventions over and above self-management and so an assumption around the proportion of incidence that would be chronic in nature was required (40%). Based on NICE guidance [58], Leeds MSK service (personal communication) and a CSAG Report [59] we estimated what proportion of a chronic population would have received each of the three identified interventions over the period of interest. For structured exercise programmes the proportion was estimated to be 5% in 1994 and 20% by 2013; for manual therapy, these figures were 3% and 20%, and for combined physical and psychological therapy 0.3% and 2.5%, respectively.

Data on per patient net health gains were taken from NICE/NCCPC guidelines [58], except for manual therapy, which was taken from a cost-effectiveness analysis of the United Kingdom BEAM trial [60].

Net monetary benefit (NMB)

Table 2 shows the contribution to the total estimates of lifetime QALYs gained from the nine areas addressed, by year, based on the estimated number of new patients in which the intervention was initiated (procedural interventions are delivered in that year only, but pharmacological treatment duration varies across interventions). By far the largest contribution to the total health gain came from improved treatment of RA (40.1% of the total). Inflammatory arthritis as a whole accounts for 57.5% of the total. Osteoarthritis and osteoporosis are the next biggest areas (21.7% and 12.2%, respectively).

Table 3 shows the lifetime net costs to the NHS of new patients initiated on the treatments in question for each of these areas over the 20-year period. Again, by far the biggest costs are associated with RA alone or inflammatory arthritis taken as a whole. It is notable that developments for osteoarthritis have led to a substantial cost saving as a result of lower surgical hospital stay observed over the period. Less substantial, but significant cost savings also arose from treatment of connective tissue disorders, as a result of avoiding the costs associated with renal failure in patients with SLE. The table also reflects that the treatments for LBP that were adopted were relatively cheap and highly cost-effective.

Table 4 summarises the NMB when the QALYs have been valued at the base-case value of £25,000 and the net costs to the NHS of the intervention and its long-term sequelae have been deducted. At this value of a

Table 2 QALYs gained from key musculoskeletal disease interventions, 1994–2013

	Rheumatoid arthritis (M00–06)	Psoriatic arthritis (M07)	Juvenile idiopathic arthritis (M08–09)	Gout (M10–12)	Osteoarthritis (M15–19)	Connective tissue disorders (M30–35)	Ankylosing spondylitis (M45)	Low back pain (M54.5)	Osteoporosis (M80–82)	Total QALYs
1994	698	67	205	1009	1944	6	0	534	0	4464
1995	1181	113	346	1015	2135	7	0	619	287	5703
1996	1786	171	523	1021	2825	8	0	882	287	7504
1997	2513	241	736	1027	3520	10	0	1171	458	9675
1998	3363	322	986	1033	4219	11	0	1468	572	11,973
1999	4964	499	1270	1038	4736	19	33	1624	853	15,036
2000	6449	641	1705	1089	5249	29	33	1847	915	17,957
2001	8459	845	2288	1128	7010	46	51	2070	1749	23,644
2002	9657	956	2641	1183	9500	63	46	2267	2400	28,713
2003	11,409	1137	3146	1245	11,737	85	69	2499	3663	34,990
2004	14,474	1457	4090	1309	14,454	109	197	2795	5022	43,907
2005	13,976	1454	4053	1157	11,504	106	212	2765	5974	41,201
2006	17,450	1361	3769	1135	11,652	129	154	2747	6833	45,229
2007	19,551	1858	5321	1260	12,929	156	367	2786	7921	52,149
2008	23,730	2381	6309	1493	13,386	181	392	4288	8955	61,116
2009	28,588	2662	7756	1471	13,383	195	482	5741	10,252	70,529
2010	33,882	3071	8945	1592	14,451	219	577	6999	11,994	81,729
2011	45,401	4007	11,994	1797	14,625	169	708	8546	12,538	99,786
2012	51,586	4453	13,422	1885	14,726	125	809	8385	13,111	108,502
2013	50,407	4813	13,248	2143	15,153	128	853	8212	12,875	107,834
Total	349,523	32,507	92,753	26,032	189,136	1801	4983	68,245	106,660	871,693
Value	£8738 m	£813 m	£2,319 m	£651 m	£4728 m	£45 m	£125 m	£1706 m	£2667 m	£21,791 m

Table 3 Net costs of delivery of key musculoskeletal disease interventions, 1994–2013

	Rheumatoid arthritis (M00–06)	Psoriatic arthritis (M07)	Juvenile idiopathic arthritis (M08–09)	Gout (M10–12)	Osteoarthritis (M15–19)	Connective tissue disorders (M30–35)	Ankylosing spondylitis (M45)	Low back pain (M54.5)	Osteoporosis (M80–82)	Total net costs of delivery
1994	£2.0 m	£0.2 m	£0.6 m	£3.3 m	£3.6 m	−£0.3 m	£0.0 m	£5.1 m	£0.0 m	£14.4 m
1995	£3.3 m	£0.3 m	£1.0 m	£3.3 m	£4.2 m	−£0.4 m	£0.0 m	£6.0 m	£4.1 m	£21.9 m
1996	£5.0 m	£0.5 m	£1.5 m	£3.3 m	£10.3 m	−£0.4 m	£0.0 m	£8.5 m	£4.1 m	£32.8 m
1997	£7.0 m	£0.7 m	£2.1 m	£3.3 m	£16.4 m	−£0.5 m	£0.0 m	£11.3 m	£14.0 m	£54.3 m
1998	£9.4 m	£0.9 m	£2.8 m	£3.4 m	£22.5 m	−£0.6 m	£0.0 m	£14.2 m	£16.0 m	£68.4 m
1999	£33.8 m	£3.6 m	£3.6 m	£3.4 m	£26.5 m	−£1.0 m	£1.3 m	£15.7 m	£20.4 m	£107.4 m
2000	£38.0 m	£4.0 m	£4.8 m	£3.6 m	£13.6 m	−£1.4 m	£1.3 m	£17.8 m	£21.3 m	£103.0 m
2001	£53.7 m	£5.5 m	£8.9 m	£3.7 m	£9.1 m	−£2.3 m	£1.7 m	£20.0 m	£36.3 m	£136.6 m
2002	£53.9 m	£5.5 m	£9.1 m	£3.9 m	−£7.7 m	−£3.1 m	£1.6 m	£21.9 m	£44.2 m	£129.3 m
2003	£71.5 m	£7.2 m	£13.7 m	£4.1 m	−£36.0 m	−£4.2 m	£2.1 m	£24.2 m	£64.3 m	£146.8 m
2004	£150.6 m	£13.4 m	£38.0 m	£4.3 m	−£47.7 m	−£5.4 m	£5.1 m	£27.0 m	£87.5 m	£272.8 m
2005	£158.4 m	£15.2 m	£44.9 m	£3.8 m	−£117.8 m	−£5.3 m	£5.5 m	£26.8 m	£108.5 m	£239.9 m
2006	£211.6 m	£11.0 m	£26.2 m	£3.7 m	−£155.2 m	−£6.4 m	£4.5 m	£26.6 m	£122.7 m	£244.7 m
2007	£275.9 m	£21.1 m	£67.8 m	£4.1 m	−£205.5 m	−£7.7 m	£9.3 m	£27.0 m	£140.4 m	£332.4 m
2008	£308.4 m	£27.0 m	£69.4 m	£4.9 m	−£234.4 m	−£9.0 m	£10.2 m	£39.9 m	£157.8 m	£374.3 m
2009	£376.8 m	£27.4 m	£90.4 m	£4.8 m	−£244.1 m	−£9.7 m	£12.5 m	£52.3 m	£176.8 m	£487.3 m
2010	£456.5 m	£31.3 m	£102.9 m	£5.3 m	−£283.0 m	−£10.9 m	£14.9 m	£63.0 m	£202.0 m	£582.0 m
2011	£594.1 m	£37.7 m	£133.5 m	£6.0 m	−£331.7 m	−£8.4 m	£17.9 m	£76.3 m	£211.7 m	£737.2 m
2012	£684.2 m	£41.0 m	£148.2 m	£6.3 m	−£343.8 m	−£6.2 m	£20.3 m	£74.8 m	£220.9 m	£845.6 m
2013	£683.9 m	£53.8 m	£154.3 m	£7.3 m	−£367.6 m	−£6.4 m	£21.6 m	£73.3 m	£215.8 m	£836.0 m
Total	£4178.2 m	£307.3 m	£923.5 m	£85.6 m	−£2268.3 m	−£89.4 m	£129.9 m	£631.5 m	£1868.9 m	£5767.2 m

QALY, all areas except treatments for ankylosing spondylitis show a positive NMB. Osteoarthritis is the single area with the largest NMB (approximately 43.7% of the total), although inflammatory arthritis as a whole accounts for a similar proportion. Within that total, however, RA accounts for 28.5%. It contributes to NMB to a lesser extent than to QALYs because the new DMARDs have generally been priced to be just acceptable to NICE at the upper end of its £20,000–30,000 'threshold'. Indeed, the new DMARDs for RA as a whole produced a negative NMB, but this was offset by large net health gains from the shift towards early, aggressive combination therapy. The overall annual figures for monetised QALYs, net cost of delivery and NMB of key MSD interventions 1994–2013 are shown Fig. 4. Additional file 2 provides details of the breakdown of estimated numbers of patients for each intervention in each year and the related QALY estimates.

Estimating the elapsed time

Our estimate of the elapsed time between research funding and health gain was based primarily on analysis of the references cited on clinical guidelines. As illustrated in Fig. 5, the mean age of the 3237 cited papers extracted from the 22 guidelines was 9 years. The median age was 7 years, with an interquartile range of 7 (4–11) years. To produce an estimate of the total elapsed time between investment and return, as required for this study, we added on to this value estimates for (1) the time between the awarding of funding and publication, and (2) the time between recommendation and realisation of health gain in clinical practice. Using the same approach as in our previous studies, we estimated these two periods to total approximately 7 years (3 years for the period between funding and publication and 4 years between recommendation and health gain). This gave a best estimated elapsed time between spending on research and health gain of 16 years. We looked at alternative approaches to estimate the knowledge cycle time, such as only including the NICE and Scottish Intercollegiate Guideline Network guidelines (mean 9 years, median 7 years) and looking at only the main osteoarthritis and RA guidelines, which are the conditions from which the largest health improvements stem (mean 10 years, median 9 years) Additional file 3 provides details of the guideline analysis.

Table 4 Net monetary benefit from key musculoskeletal disease interventions, 1994–2013

	Rheumatoid arthritis (M00–06)	Psoriatic arthritis (M07)	Juvenile idiopathic arthritis (M08–09)	Gout (M10–12)	Osteoarthritis (M15–19)	Connective tissue disorders (M30–35)	Ankylosing spondylitis (M45)	Low back pain (M54.5)	Osteoporosis (M80–82)	Total net monetary benefit
1994	£15.5 m	£1.5 m	£4.5 m	£21.9 m	£45.0 m	£0.4 m	£0.0 m	£8.2 m	£0.0 m	£97.2 m
1995	£26.2 m	£2.5 m	£7.7 m	£22.1 m	£49.1 m	£0.5 m	£0.0 m	£9.5 m	£3.0 m	£120.7 m
1996	£39.6 m	£3.8 m	£11.6 m	£22.2 m	£60.3 m	£0.6 m	£0.0 m	£13.6 m	£3.0 m	£154.8 m
1997	£55.8 m	£5.3 m	£16.4 m	£22.3 m	£71.6 m	£0.7 m	£0.0 m	£18.0 m	-£2.6 m	£187.6 m
1998	£74.7 m	£7.1 m	£21.9 m	£22.4 m	£83.0 m	£0.8 m	£0.0 m	£22.5 m	-£1.6 m	£230.9 m
1999	£90.3 m	£8.8 m	£28.2 m	£22.6 m	£91.9 m	£1.4 m	−£0.5 m	£24.9 m	£0.9 m	£268.5 m
2000	£123.2 m	£12.0 m	£37.9 m	£23.7 m	£117.6 m	£2.1 m	−£0.5 m	£28.3 m	£1.6 m	£345.9 m
2001	£157.8 m	£15.6 m	£48.3 m	£24.5 m	£166.2 m	£3.4 m	−£0.5 m	£31.7 m	£7.5 m	£454.5 m
2002	£187.5 m	£18.4 m	£56.9 m	£25.7 m	£245.2 m	£4.7 m	−£0.5 m	£34.8 m	£15.8 m	£588.5 m
2003	£213.7 m	£21.3 m	£65.0 m	£27.1 m	£329.4 m	£6.3 m	−£0.4 m	£38.3 m	£27.3 m	£728.0 m
2004	£211.2 m	£23.0 m	£64.3 m	£28.5 m	£409.1 m	£8.1 m	−£0.2 m	£42.8 m	£38.0 m	£824.9 m
2005	£191.0 m	£21.1 m	£56.4 m	£25.1 m	£405.4 m	£7.9 m	−£0.2 m	£42.4 m	£40.9 m	£790.1 m
2006	£224.6 m	£23.1 m	£68.0 m	£24.7 m	£446.5 m	£9.7 m	−£0.6 m	£42.1 m	£48.1 m	£886.1 m
2007	£212.8 m	£25.4 m	£65.2 m	£27.4 m	£528.7 m	£11.6 m	−£0.2 m	£42.7 m	£57.6 m	£971.3 m
2008	£284.8 m	£32.5 m	£88.4 m	£32.5 m	£569.0 m	£13.5 m	−£0.4 m	£67.3 m	£66.1 m	£1153.6 m
2009	£337.9 m	£39.1 m	£103.5 m	£32.0 m	£578.6 m	£14.5 m	−£0.4 m	£91.2 m	£79.5 m	£1276.0 m
2010	£390.5 m	£45.4 m	£120.7 m	£34.5 m	£644.3 m	£16.3 m	−£0.4 m	£112.0 m	£97.8 m	£1461.2 m
2011	£540.9 m	£62.5 m	£166.3 m	£38.9 m	£697.3 m	£12.6 m	−£0.2 m	£137.4 m	£101.8 m	£1757.4 m
2012	£605.4 m	£70.3 m	£187.4 m	£40.8 m	£712.0 m	£9.3 m	£0.0 m	£134.8 m	£106.9 m	£1867.0 m
2013	£576.3 m	£66.5 m	£176.8 m	£46.3 m	£746.4 m	£9.6 m	−£0.2 m	£132.0 m	£106.1 m	£1859.8 m
Total	£4559.9 m	£505.3 m	£1395.4 m	£565.2 m	£6996.8 m	£134.4 m	−£5.3 m	£1074.6 m	£797.6 m	£16,023.8 m

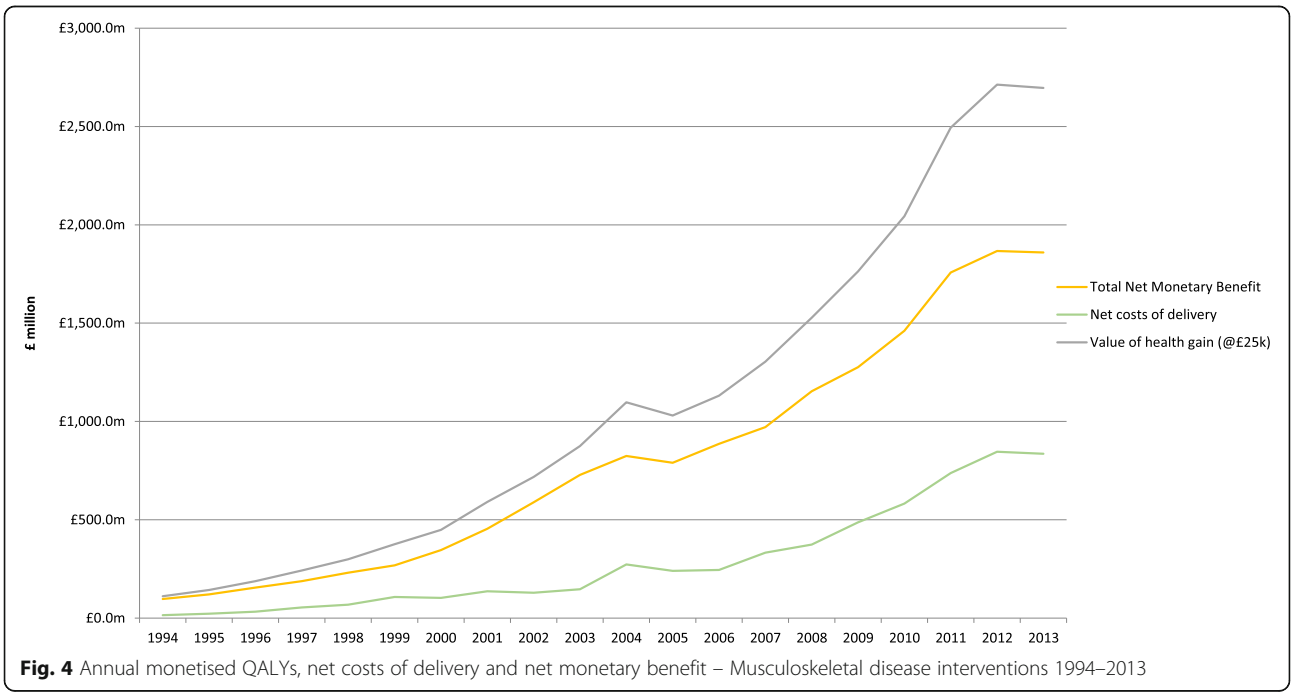


Fig. 4 Annual monetised QALYs, net costs of delivery and net monetary benefit – Musculoskeletal disease interventions 1994–2013

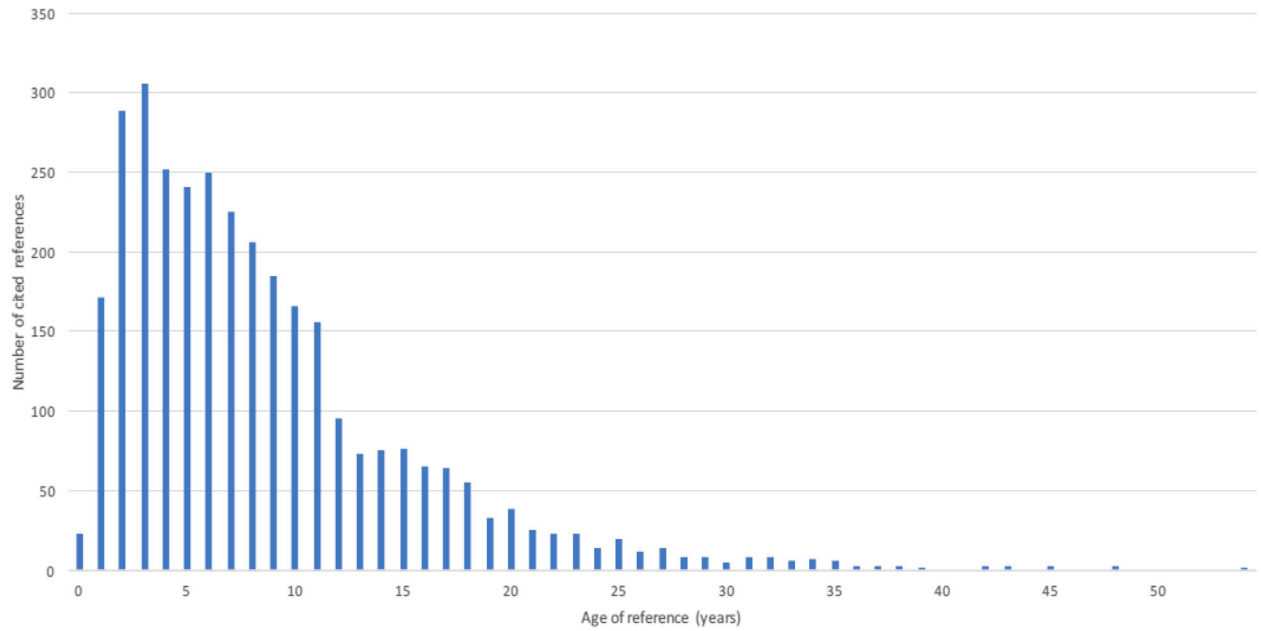


Fig. 5 Elapsed time of the cited papers extracted from guidelines

Estimating the proportion of health gains that can be attributed to United Kingdom research

The estimate of the proportion of the health gain that can be attributed to United Kingdom research was also based primarily on the analysis of cited references on clinical guidelines. A total of 2762 publications were analysed. The overall percentage across all guidelines, using full counting[6] as for the previous studies, was 30%, which forms our central estimate, but as shown in Table 5, this differed substantially between specific guidelines. We also produced overall estimates using fractional counting[7] and the reprint address,[8] which gave an attribution to the United Kingdom of 25% and 24%, respectively.

To produce a range of values for the sensitivity analysis, we can consider the potential sources of uncertainty in these estimates. We identify two likely sources of error.

Firstly, we assume, for the purposes of our analysis, that the proportion of research conducted in the United Kingdom corresponds to the proportion supported by United Kingdom (charitable or public) funding. However, United Kingdom authors may receive funding from the United Kingdom or overseas industry or from other non-United Kingdom sources (notably the European Commission, but also other international funders). Equally, United Kingdom funders may fund researchers overseas, but we expect this to be limited in this field, and in most cases this is likely to be in collaboration with at least one United Kingdom author, in which case the full counting model would capture the resulting publications. For the purposes of our model, we assume flows of funding into and out of the United Kingdom to be equal. Considering industry funding, it may be that some of the papers with a United Kingdom address are industry funded (including non-United Kingdom industry), and as such should be excluded from the number of United Kingdom papers for our estimate of attribution to the United Kingdom. We expect this proportion to be small, but this is clearly an issue which warrants further investigation.

Secondly, there is uncertainty around the relative contribution of funding associated with each author (and hence country) listed on each paper. The three bibliometric methods all estimate this differently. Using full counting effectively assumes that the United Kingdom contributes all the funding for any paper which has a United Kingdom author (and does the same for any other countries on the same paper). This is likely to overestimate the United Kingdom contribution. Fractional counting at the author level assumes an equal contribution of funding from each author on a paper (from the country in which they are based). With reprint addresses, the assumption is that all funding comes from the country in which the corresponding author is based. For the last two approaches, it is not clear whether they are likely to give an under- or overestimation of the proportion of funding from the United Kingdom. For consistency with previous studies, we have used the full counting approach for our central estimate.

Based on this analysis, we conclude that it is unlikely that the United Kingdom contribution is higher than our estimate from full counting of 30% (as used in

Table 5 Proportion of publications from the United Kingdom for all guidelines included in the analysis

Guideline	United Kingdom papers	Total papers	% United Kingdom
British Association/College of Occupational Therapists – Hand and wrist orthoses for adults with rheumatological conditions: practice guideline for occupational therapists (evidence)	5	25	20%
British Association/College of Occupational Therapists – Hand and wrist orthoses for adults with rheumatological conditions: practice guideline for occupational therapists (supplementary)	10	20	50%
British Association/College of Occupational Therapists – Occupational therapy for adults undergoing total hip replacement: practice guideline for occupational therapists (evidence)	9	30	30%
British Association/College of Occupational Therapists – Occupational therapy for adults undergoing total hip replacement: practice guideline for occupational therapists (supplementary)	1	4	25%
British Pain Society – The assessment of pain in older people	6	63	10%
British Pain Society – Guidelines for pain management programmes for adults	21	55	38%
British Society for Rheumatology – British Society for Rheumatology and IASP Musculoskeletal Pain Taskforce guidelines for the integrated management of musculoskeletal pain symptoms (IMMsPS)	87	304	29%
British Society for Rheumatology – BSR guidelines on standards of care for persons with rheumatoid arthritis	1	1	100%
National Institute for Health and Care Excellence – Hip fracture (CG.124)	73	254	29%
National Institute for Health and Care Excellence – Osteoporosis (CG.146)	32	71	45%
National Institute for Health and Care Excellence – Osteoarthritis (CG.177)	102	416	25%
National Institute for Health and Care Excellence – Rheumatoid arthritis in adults (CG.79)	102	337	30%
National Institute for Health and Care Excellence – Low back pain in adults (CG.88)	31	111	28%
National Osteoporosis Foundation – Clinician's guide to the prevention and treatment of osteoporosis	33	90	37%
National Osteoporosis Guideline Group – Guideline for the diagnosis and management of osteoporosis in postmenopausal women and men from the age of 50 years in the United Kingdom	2	2	100%
National Osteoporosis Guideline Group – Osteoporosis: clinical guideline for prevention and treatment: executive summary	22	36	61%
National Osteoporosis Society – Vitamin D and bone health: a practical clinical guideline for patient management	25	66	38%
Royal College of Nursing – Administering subcutaneous methotrexate for inflammatory arthritis	14	56	25%
Royal College of Physicians – Pain: complex regional pain syndrome	31	96	32%
Royal College of Physicians – Upper limb disorders: occupational aspects of management 2009	13	52	25%
Scottish Intercollegiate Guidelines Network – Management of hip fracture in older people (CG.111)	41	102	40%
Scottish Intercollegiate Guidelines Network – Management of early rheumatoid arthritis (CG.123)	23	83	28%
Scottish Intercollegiate Guidelines Network – Management of chronic pain (CG.136)	56	171	33%
Scottish Intercollegiate Guidelines Network – Management of osteoporosis and the prevention of fragility fractures (CG.142)	92	317	29%
TOTAL	832	2762	30%

previous studies). However, it may be lower than our lowest estimate using the reprint addresses of 24%, considering the other potential sources of funding available to United Kingdom-based researchers. Taking this into account, we used lower and upper bounds of 20% and 30% for the sensitivity analysis.

Estimating the IRR from musculoskeletal disease research

Our estimates of the NMB produced by year (summarised in Table 4) at a base-case value of a QALY of £25,000 were then related to our best estimates of public and charitable spend by year on MSD research (summarised in Fig. 3) and expressed as an IRR. Calculation of the IRR incorporates our best estimates of the average elapsed time between research spending and use of the intervention (16 years) and of the proportion of the NMB that could be attributable to United Kingdom research (30%). This gives a base-case estimate of an IRR of 6.8%.

As is evident from the methods used, there is inevitably considerable uncertainty around the values of all our estimates. Table 6 presents a series of one-way

Table 6 Internal rate of return: one way sensitivity analyses

	IRR
Best Estimate	6.8%
Low research spend (£12 m)	7.6%
High research spend (£16 m)	6.0%
Omit length of stay reduction	5.5%
QALY £13 k	0.8%
QALY £20 k	5.0%
QALY £30 k	8.1%
QALY £60 k	12.9%
Long lag (20 years)	5.5%
Short lag (11 years)	8.1%
Low attribution to United Kingdom (20%)	4.5%
NMB −25%	5.1%
NMB +25%	8.0%
Royalty payments to public/charitable funders	7.0%
Total royalty payments	7.2%

IRR internal rate of return, *QALY* quality adjusted life years, *NMB* net monetary benefit

sensitivity analyses to illustrate the effects of some of the main areas of uncertainty – all changes have predictable effects. Despite the detail of our estimation process there is considerable uncertainty in the NMB; we present the implications for the IRR of an arbitrary but plausible range of –25% and +25% around our estimate to reflect this. We also present the IRR omitting the cost-savings from reduction in length of stay for hip and knee replacements (see Discussion). The impact of taking into account the royalty payments arising from anti-TNF research increased the IRR by 0.2 percentage points taking our lower figures (possible underestimate) and by 0.4 percentage points using our higher figures (likely overestimate).

The IRR predictably decreases with increased estimates of research funding and elapsed time and, as far as is explored, all the variables in the one-way sensitivity analyses show a positive rate of return. However, in combination, they could of course have produced a wider range of estimates for the IRR. Table 6 shows that, inevitably, the IRR is most sensitive to the range of plausible values that can be placed on the value of a QALY. At an opportunity cost in the NHS of £13,000, the IRR falls to 0.80%, whilst at a societal valuation of £60,000 the IRR is 12.9%.

Discussion

In this paper, we have estimated the economic returns from public and charitable funding of MSD-related research in the United Kingdom. Expressed in 2013 prices, total expenditure on MSD-related research was £1.4 billion for the period (1978–1997) that was used to estimate the rate of return. Over the period 1994–2013, the key interventions we analysed produced 871,000 QALYs with a NMB of £16 billion, allowing for the net NHS costs resulting from them and valuing a QALY at £25,000. The proportion of benefit attributable to United Kingdom research was 30% and the elapsed time between funding and impact of MSD treatments was 16 years. Our best estimate of the IRR from MSD-related research was 7%, very similar to the 9% for CVD research and 10% for cancer research (Table 7). When combined with previous estimates of the broader

Table 7 Comparison of key results with previous studies

	MSD	Cancer	CVD
Average annual research investment (for years of data used in IRR calculation as reported in source publications, using different time period for calculating constant prices and therefore not suitable for comparisons)	£70 m (1978–1997, in constant 2013–2014 prices)	£266 m (1976–1995, in constant 2011–2012 prices)	£111 m (1975–1998, in constant 2005–2006 prices)
Average annual research investment (rebased in same constant prices for comparative purposes)	£70 m (1978–1997, in constant 2013–2014 prices)	£290 m (1976–1995, in constant 2013–2014 prices)	£133 m (1975–1998, in constant 2013–2014 prices)
Elapsed time (between spending on research and health gain)	16 years	15 years	17 years
Attribution (proportion of papers that include a United Kingdom address from the papers cited on guidelines)	30%	17%	17%
Average NMB (for years of data used in IRR calculation as reported in source publications, but using different time period for calculating constant prices therefore not suitable for comparisons)	£801 m (1994–2013, in constant 2013–2014 prices)	£6223 m (1991–2010, in constant 2011–2012 prices)	£2949 m (1992–2005, in constant 2005–2006 prices)
Average NMB (rebased in same constant prices for comparative purposes)	£801 m (1994–2013, in constant 2013–2014 prices)	£6458 m (1991–2010, in constant 2013–2014 prices)	£3559 m (1992–2005, in constant 2013–2014 prices)
IRR (health gain)	7%	10%	9%

CVD cardiovascular disease, *IRR* internal rate of return, *MSD* musculoskeletal disease, *NMB* net monetary benefit

economic (or 'spillover') benefits of biomedical and health research in the United Kingdom of 17% [4], the total rate of return is approximately 24–27%.

In this study, we have also further tested the bottom-up methodological approach developed in the original 'Medical Research: What's it worth?' study [2]. The application of this method to a further disease area that is different to CVD and cancer – particularly in terms of the chronic nature of MSD and the predominantly quality of life gains of the benefit of the interventions – confirms the generalisability of our approach to estimate the economic returns from research; that is not to say that it is without limitations. We have organised the discussion on limitations and caveats by first looking at the key conceptual issues with the methodological approach, followed by issues related to data availability and quality, then an examination of a set of issues related to MSD research, and, finally, a set of key caveats on what this work demonstrates and what it does not.

Key conceptual assumptions inherent to methodological approach

In estimating the economic returns from MSD-related research (and indeed in the previous studies looking at CVD and cancer research) various key assumptions are made that are intrinsic to the conceptual approach adopted:

- **The total NMB for the interventions not covered is assumed to be zero.** Our estimate assumes that any other MSD interventions introduced or widely adopted during the period in question not included in the analysis have, in aggregate, no effect on the NMB, that is, their NMB is equivalent to zero. Put another way we assume that, for any omitted interventions, the monetised value of the health benefit is equal to the cost of delivering the benefit. This seems a reasonable assumption as there may be interventions where the cost of delivery outweighs the value of the benefit and others where the value of the benefit outweighs the costs of delivery. Without analysing all these other interventions, it would be wrong to speculate on the balance of these effects and therefore they are assumed to cancel each other out. However, as discussed below, the likelihood of this assumption being correct will vary as the value of the QALY is decreased or increased in the sensitivity analysis.
- **The total net flow of knowledge between disciplines is assumed to be zero.** We know that the relationship between research discipline and impact is 'many-to-many' [61], that is research from a specific discipline will contribute to multiple types of impact and a specific impact is often made up of contributions from multiple research disciplines. In the context of the current study, it is likely, for example, that MSD-related research benefits from, say, cancer research and vice versa. We therefore assume that the flow of knowledge is the same in to as it is out of different research fields, in effect cancelling each other out.
- **The definition of MSD-related research used by the research funders captures basic research.** We know this is the case for ARUK and the other disease-specific medical research charities as all their research funding is included in the analysis. For the FCs and the DH/NHS this would not be an issue as estimates for their MSD-related research funding were derived by applying the 'activity index', which would include basic research. However, for the MRC and WT this could be an issue. For the MRC, we relied on the funder's classification and used the broader of two definitions so that we would deliberately err on the side of caution by taking the higher level of R&D spend. For the WT, we had to rely on search terms and in scanning research grant titles we were reassured that fundamental research was included.
- **The cost of private sector R&D is covered in the net NHS costs of the interventions.** We assume that the costs of private sector R&D (i.e. non-public and non-charitable research expenditure) are accounted for when we net off the NHS costs for an intervention. This assumption holds for purely commercial research as, say, a pharmaceutical company will include the cost of their R&D investments in the price of a drug. It may be that companies invest in 'non-commercial' activities, such as public–private partnerships or precompetitive consortia, and in effect are subsidising the public sector research in doing so. However, even in this case (which is probably at the margins of total R&D investments) it is unlikely that the private sector is doing so in isolation of commercial considerations and it will recoup such costs through its sales revenues.
- **All health gains arise from specific patient interventions.** We assume that all health gains arise from, and are captured in, our estimates of the health gain from specific patient interventions. We recognise that broader service changes, such as the adoption of fracture clinics, or improvements in diagnosis are important but assume that they lead to patients receiving timely and appropriate interventions for which we estimate the QALYs gained. There is a possibility that we are failing to net off the full costs of such developments in service delivery if the cost-

effectiveness evidence we use for such interventions fails to reflect the full cost of the service delivery associated with them.

- **We have assumed there is a causal relationship between research and health gains.** Our analysis relies on the assumption that the health benefits would not have occurred without the evidence from the medical research. Our bottom-up approach has the advantage that, for the individual interventions, there is causality as demonstrated through their formal clinical trials. Additionally, in this disease area we do not have the uncertainty as to any causal factors, other than medical research, that may have led in part to a reduction in smoking. Furthermore, in previous research on MSD, we used a case study approach that clearly demonstrated causality for the small number of interventions examined [62, 63].

Uncertainties relating to key parameters

There remain a number of uncertainties with the bottom-up approach that relate to the nature, quality and availability of data that are relevant to this examination of MSD research, and were also the case for CVD and cancer research. Reducing the uncertainty with these data issues would improve the robustness of the IRR estimate and thus, in part, could inform future research avenues.

- **The monetary value of a QALY.** As noted earlier in the paper, there is ongoing debate as to the appropriate value of a QALY. Our base-case assumption of £25,000 is consistent with our analysis of the returns to CVD and cancer research, and reflects the mid-point in the range of values (of £20,000 to £30,000) cited as normal criteria for acceptance of interventions by NICE [24]. However, as highlighted above, if the QALY is valued either at a lower (e.g. £13,000) or higher (e.g. £60,000) level then this could affect our core assumption that the total NMB for any new interventions not covered is assumed to be zero. If QALYs are valued at £60,000 then more interventions are likely to have a positive NMB among those not looked at, meaning that we are underestimating the rate of return. Conversely, if the QALY is valued at £13,000, then more interventions in those not looked at are likely to have a negative NMB, meaning that we are over estimating the rate of return.
- **Estimates of the elapsed time are hard to determine.** As with the previous studies, bibliometric analysis of clinical guidelines was used to estimate the time between research investment and health gain. The advantage of this approach is that it provides empirical estimates, but it is also, inevitably, a gross simplification of a complex and varied process, as we have discussed elsewhere [64, 65]. The estimate of the elapsed time is in accordance with other estimates using different approaches as reviewed by Morris et al. [66], but is still a crude proxy and is an area that would benefit from further research.
- **Estimates for the rate of attribution are very hard to determine.** Like the estimate of the elapsed time, bibliometric analysis of clinical guidelines was used to estimate the proportion of the United Kingdom health gains that can be related to United Kingdom public and charitable research funding. However, the estimate of the attribution rate is harder to validate than the elapsed time and is thus more contestable. It is also becoming increasingly difficult to define, given the steady increase in international collaboration in research observed in recent decades [67]. Identifying any one country's contribution without a qualitative assessment of the research itself can only provide an uncertain estimate, but one that we believe is likely to be more robust at the aggregate level of an entire research field. Biases in coverage of bibliometric databases, particularly as regards languages other than English, should also be noted. An attribution rate of at least 7–9% would be expected given that the United Kingdom contributes approximately 7–9% of biomedical and health research outputs [68]. One could also argue that the rate would be somewhat higher than this given that the local healthcare context is likely to drive the need for locally relevant studies. In the previous two studies, the attribution rate was 17% for both CVD and cancer (Table 7), which, given the above logic, felt defensible. However, an attribution rate of 30% for MSD seems high and was at the top end of the estimates we generated using different bibliometric methods. Thus, and although we used 30% for consistency with the previous studies, we did not include a higher upper bound estimate in the sensitivity analysis. It may also be that a proportion of papers cited on the clinical guidelines are solely private sector-funded and thus overstate the attribution rate to publically funded research (a scan of the references suggests that approximately 10% of United Kingdom papers could be solely industry funded). Either way a clear priority for future research would be to further examine how you measure how much of the United Kingdom health gain you can attribute to United Kingdom public and charitable research, and to validate or otherwise the guideline methodology.
- **Missing funding data.** Historical data on research funding expenditure was incomplete, meaning that

we had to make a number of assumptions to account for missing data. These assumptions erred on the side of caution and were tested in the sensitivity analysis of the IRR. As we have previously noted [2], if research funders wish to carry on with this type of analysis, the continued use of standard systems of research classification such as the HRCS will be important.

Key issues particular to MSD

There are a number of specific issues that relate to the assessment of MSD research, as indicated below.

- **Quality of data.** We had expected that there might be greater problems in identifying cost-effectiveness data in MSD given that the outcomes of interventions are principally improvements in quality of life. In practice, there was relatively good data on the cost-effectiveness and usage of new drugs, reflecting that many had been subject to NICE appraisals. By contrast, there was much less cost-effectiveness data and very poor data on provision and usage for some interventions such as those for back pain. In part, this reflects the complexity and variability of the physical therapies potentially provided to multiple groups of MSD patients. There are issues about the generalisability of clinical trials and associated cost-effectiveness studies and an absence of consistent routine methods of data collection on their usage.
- **Pricing of new pharmacological interventions.** One of the major therapeutic developments in the period was the advent of biologic DMARDs including anti-TNFs. Our analysis shows that these made a substantial contribution to the QALYs gained, but as most were priced to try to meet NICE's cost-effectiveness 'threshold' they contributed rather little to our estimates of NMB. However, as biosimilars now become available these drugs will, in future, if not superseded by other new interventions, contribute more strongly to any estimate of NMB. By contrast, early aggressive treatment with generic methotrexate provides both QALYs and a high NMB.
- **Importance of reductions in length of stay.** Our clinical experts emphasised the importance of changes in practice during the relevant period that had reduced lengths of stay and hence costs for some key procedures. They noted that, whilst desire for the reduction may have been driven by cost considerations, the change in practice was supported by research evidence showing no reduction in health benefit [69–71]. Review of the data showed that the change was marked and we estimated the cost savings in the case of hip and knee replacements. However, we are aware that there may have been some similarly marked changes that we did not quantify in previous studies. Whilst we do not believe they would have been so significant in the case of CVD or cancer we provide a sensitivity analysis to show the IRR for MSD if these cost savings are excluded.
- **Other interventions that we might have included.** Some advances in the treatment of musculoskeletal connective tissue disorders have occurred over the period of interest, but data on their cost-effectiveness is limited. For example, there was a lack of cost-effectiveness data for other immunosuppressant therapies in SLE and scleroderma, notably cyclophosphamide and intravenous immunoglobulin for the treatment of dermatomyositis. The net health gains associated with rituximab use in SLE were also not quantified in the model due to a lack of data. Similarly, we were unable to characterise the cost-effectiveness and to find appropriate data on specific treatments for soft-tissue musculoskeletal pain (M60–M79). As noted above, for any area we were not able to analyse specifically, our methods implicitly assume that any benefits from treatment were directly offset by their costs of delivery (i.e. the NMB is 0). However, we are confident that our analysis directly captures most of the significant advances in the field that have produced important health or cost effects when viewed at a population level, and unlike our analysis of cancer, because the MSD field is smaller, we did not have to prioritise and effectively ignore some potentially important areas.
- **The treatment of royalty payments.** The particular circumstances of the commercialisation of anti-TNF research led us to consider for the first time in this study the impact on the IRR of the significant royalty payments, although we were unable to establish the precise total magnitude of the royalty payments that were returned to publicly funded medical research. The impact was not negligible although they did not significantly change the order of magnitude of the IRR. We are not aware of royalties of a similar relative magnitude in the case of our previous studies on CVD and cancer, but clearly it is an issue that needs to be considered and refined in future studies.

Key limitations and caveats to the 'bottom-up' approach for assessing economic returns

There are three key caveats that are fundamental to appropriately using the results presented in this and the previous papers in assessing the economic rates of return for MSD, cancer and CVD research.

- **We have assessed past performance, not predicted future performance.** In all three studies, we have estimated the rate of return based on past investments and therefore our results cannot be a guarantee of future returns – medical research does not advance in a smooth and linear fashion. This is a crucial caveat when using these results to advocate the need for future research spending.
- **We have assessed the average rate of return, not the marginal rate of return.** From the analysis, we are not able to say whether the rate of return would have been different if research spending had been higher or lower, or whether at the margin the returns to research investment are increasing or decreasing. Assessing the marginal rate of return is of clear interest to policy-makers and this is a topic that warrants further research attention.
- **Our estimates should not be used to make comparisons between disease area or intervention.** Given the inherent assumptions and uncertainty in our approach we strongly counsel against making comparisons between the three disease areas we have examined or specific interventions within those disease areas. Taking the three studies together, we believe it is appropriate to say the IRR arising from health gains to the United Kingdom, from United Kingdom research, is between 7% and 10%.

It is of course impossible to know whether our estimates of the return are accurate in the absence of an observation of a control to provide the counterfactual. The likely validity of our estimates can only be judged through the reasonableness of the many assumptions about which we have tried to be entirely transparent. Should readers have more or less confidence in them because of the similarity between the estimates for IRR arising from health gains across the three studies? The results are indeed remarkably comparable (Table 7). This can either be interpreted positively as some form of internal validation of plausible magnitudes, or negatively to suggest that something inherent in the methodology leads to this similar outcome. We are not aware of any aspect of our methods that suggests this is a methodological artefact. Rather, we draw some comfort from the fact that the similarity arises despite the inputs (research expenditure) and outcomes (NMB) being very different between the studies and for separate groups of interventions within each of them. The fact that the elapsed times are similar has a degree of face validity, as does the relatively high level of attribution for MSD, which, as already noted, given its chronic nature, is more likely to be influenced by local contextual research.

Conclusion

The public, both as taxpayers and charity donors, invest a significant amount of money into biomedical and health research each year. Understanding the economic impact of this investment provides accountability, helps secure future research investments and increases our understanding of how research is effectively translated into health improvements. In a series of studies looking at the net value of improved health outcomes, in CVD, cancer and MSD we have demonstrated an IRR of between 7% and 10%. When we include the 17% return for the broader economic or 'spillover' impact this rises to between 24% and 27%. The results suggest that, despite the uncertainties around the methods and estimates, the historical returns in terms of NMB of the health gains derived in the United Kingdom from public and charitably funded biomedical and health research are substantial and justify the investments made.

Endnotes

[1]The Health Research Classification System (HRCS) is a two-dimensional framework. Codes from both HRCS dimensions are applied when classifying; one dimension, the Health Categories, is used to classify the type of health or disease being studied. There are 21 categories encompassing all diseases, conditions and areas of health. The other dimension, the Research Activity Codes, classifies the type of research activity being undertaken (from basic to applied). There are 48 codes divided into eight groups. See http://www.hrcsonline.net/rac/rac for more details. The data cited are on DALY rates from 2012 and research spend data from 2014.

[2]Figures given between 1992 and 1993 show the UFC grants after the funds were transferred from the UFC to the Research Councils.

[3]Prepared by the HEFCE Analytical Services Group on 13 February 2008. Figures for 1993–1994 to 1996–1997 were adjusted to include an estimate of funds for research capital that were rolled into mainstream quality-related research (QR) from 1997–1998 onwards. No adjustments have been made to counterbalance the effect of the phased transfer of funds to the United Kingdom Research Councils in 1992–1993, 1993–1994 and 1994–1995. Figures for Units of Assessment (UoA) 12: Biochemistry, which ceased to exist in the 2001 Research Assessment Exercise, have been rolled into UoA 14: Biological Sciences. Figures exclude funds for the supervision of students on research degree programmes, London weighting, and all other relatively minor elements of research funding. For 2006–2007, 2007–2008 and 2008–2009, the QR charity support element has

been added to mainstream QR funds to reflect the change in the way research income from charities is used in the calculation of funding. Figures contain relatively minor grant adjustments made to the database after the initial grant allocations announced in March each year.

[4]Figures for 2010–2011 do not reflect the 1.7% retrospective reduction announced in HEFCE Circular Letter 05/11 and applied at institution level. In the calculations that include quality-related research (QR) charity support funding (below), the proportion of QR charities support funding attributed to biomedical Units of Assessment in 2011–2012 has been applied to 2012–2013.

[5]The Centre for Science and Technology Studies (CWTS) maintains a bibliometric database of all scientific publications (including health and biomedical research) for the period 1981 to 2016. This dataset is based on the journals and serials processed for the Core Collection version of the Web of Science database, including the Science Citation Index Expanded (SCI-E) and associated citation indices, the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCI). This database is operated for bibliometric purposes in service contracts under a License Agreement with Clarivate Analytics. See: https://www.cwts.nl/ for more information.

[6]Full counting is an approach in which a paper which has authors from multiple countries will be attributed in full to each of these countries. Therefore, the estimate of 30% indicates that the United Kingdom contributed to 30% of the papers analysed. However, there will be contributors from other countries to many of these papers. This approach is used for the central estimate as it was used in the previous studies, to aid comparison.

[7]In fractional counting, attribution of a paper is shared between the countries of origin of the various authors. For example, if a paper has one author from the United Kingdom and two from the United States, the United Kingdom will receive attribution for one third of a paper, and the other two thirds will go to the United States.

[8]This is an alternative approach where papers are attributed to the country of origin of the corresponding author. The logic is that it is likely that the institution of the corresponding author held a significant proportion of the funding for the work and made a significant contribution to the work.

Additional files

Additional file 1: Appendix 1. Funding data. (XLSX 21 kb)

Additional file 2: Appendix 2. Health gain (i.e. net monetary benefit) data. (XLSX 61 kb)

Additional file 3: Appendix 3. Guideline data. (XLSX 16 kb)

Abbreviations

AMRC: Association of Medical Research Charities; Anti-TNF: anti-tumour necrosis factor drugs; ARUK: Arthritis Research UK; bDMARDs: biologic disease-modifying anti rheumatic drugs; cDMARDs: conventional disease-modifying anti rheumatic drugs; CPRD: Clinical Practice Research Datalink; CVD: cardiovascular disease; CWTS: Centre for Science and Technology Studies; DALYs: disability adjusted life years; DH: Department of Health; DMARDs: disease-modifying anti rheumatic drugs; EACT: early and aggressive combination therapy; FC: Funding Councils; GP: general practitioner; HPE: Hospital Prescribing England; HRCS: Health Research Classification System; ICD: International Statistical Classification of Diseases and Related Health Problems; IRR: internal rate of return; JIA: juvenile idiopathic arthritis; LBP: low back pain; MRC: Medical Research Council; MSD: musculoskeletal disease; NCC: National Collaborative Centre; NHS: National Health Service; NIC: net ingredient cost; NICE: National Institute for Health and Care Excellence; NMB: net monetary benefits; PCA: prescription cost analysis; QALY: quality adjusted life years; RA: rheumatoid arthritis; R&D: research and developments; SLE: systemic lupus erythematosus; WT: Wellcome Trust.

Acknowledgements

We are grateful for financial support from the Wellcome Trust, Arthritis Research UK, the Medical Research Council, the National Institute for Health Research and the Academy of Medical Sciences. The funders had no influence in the study design, data collection and analysis, and decision to publish, or preparation of the manuscript. Through membership of the Advisory Board the funders did provide ongoing suggestions and advice throughout the study.

We would like to thank members of the Advisory Board for their advice and support, particularly Liz Allen, Nicola Perrin and Louise Wren for chairing, along with David Cox, Alan Silman, Laura Boothman, Natalie Carter, Stephen Simpson, Richard Francis, Ben Bleasdale, Ian Viney, Jorge Mestre Ferrandiz and Rachel Quinn; Rodrigo Costas (CWTS, Leiden) for providing bibliometric analysis; Martin Gulliford (KCL, London) for providing a data analysis using primary care data found in CPRD; Rachel Burden (AMRC) for providing data on the top charities funding MSD in the United Kingdom; as well as individuals from the charities which provided data, including Claire Severgnini (National Osteoporosis Society), Caroline Sangan (National Osteoporosis Society), Vicki Hughes (Nuffield Foundation/Oliver Bird), Simon Gallacher (Nuffield Foundation/Oliver Bird), Kiki Syrad (Great Ormond Street Hospital), Ali Momin (Great Ormond Street Hospital), Tracy Swinfield (Action Medical Research), Arash Angadji (Orthopaedic Research United Kingdom), Martyn Coomer (Royal College of Surgeons), Sarah Allport (Dunhill Medical Trust), Mike Kimmons (British Orthopaedic Association), and Mark Bacon (Spinal Research).

We would especially like to extend our thanks to all those MSD experts that we have consulted at various stages of the project, particularly Maya Buch, Philip Conaghan, Peter Crof, Elaine Dennison, Lucy Wedderburn, who were crucial to verifying data; as well as Andrew Carr, Anthony Redmond, Mike Hurley, Paul Emery, David Marsh, Martin Underwood, Mark Layton and David D'Cruz.

To determine total incidences of lower back pain per year, primary care data was extracted and analysed from electronic records in the CPRD database with approval from the ISAC committee.

Needless to say, we are responsible for any errors, misrepresentations or inaccuracies.

Funding

This study was funded through a research grant from the Wellcome Trust, Arthritis Research UK, the Medical Research Council (MRC), the National Institute for Health Research (NIHR) and the Academy of Medical Sciences.

Availability of data and materials

The synthesised data used and generated during this study are included in the published paper and Additional files.

Authors' contributions

The project was conceived by JG and MB, and designed and executed by all the authors. EM worked as a full-time research assistant supporting all elements of the study. JG and EM worked on the funding analysis; AP, SG and EM on the guidelines analysis; and MG, MB and EM on the assessment of health gain and the calculation of the IRR. All authors were involved in synthesising and

interpreting the results. All authors contributed drafts for various parts of the paper, critically reviewing various iterations and approving the final draft submitted.

Ethics approval and consent to participate
Ethical approval for this research was obtained from King's College London through its minimal risk process (Research Ethics Number: MR/15/16-11).

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Health Economics Research Group, Brunel University London, Uxbridge, United Kingdom. [2]Policy Institute at King's, King's College London, Virginia Woolf Building, 22 Kingsway, London WC2B 6LE, United Kingdom. [3]RAND Europe, Cambridge, United Kingdom.

Received: 24 October 2017 Accepted: 12 December 2017
Published online: 10 January 2018

References

1. Røttingen J-A, Regmi S, Eide M, Young AJ, Viergever RF, Ardal C, Guzman J, Edwards D, Matlin SA, Terry RF. Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory? Lancet. 2013;382:1286–307.
2. Health Economics Research Group, Office of Health Economics, RAND Europe. Medical Research: What's it Worth? Estimating the Economic Benefits from Medical Research in the UK. London: UK Evaluation Forum; 2008. http://www.brunel.ac.uk/__data/assets/pdf_file/0008/183455/TAP825EconomicBenefitsReportFULLWeb.pdf. Accessed March 2017.
3. Glover M, Buxton M, Guthrie S, Hanney S, Pollitt A, Grant J. Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. BMC Medicine. 2014;12:99.
4. Sussex J, Feng Y, Mestre-Ferrandiz J, Pistollato M, Hafner M, Burridge P, Grant J. Quantifying the economic impact of government and charity funding of medical research on private R&D funding in the United Kingdom. BMC Medicine. 2016;14:32.
5. Buxton M, Hanney S, Jones T. Estimating the economic value to societies of the impact of health research: a critical review. Bull World Health Organ. 2004;82:733–9.
6. Raftery J, Hanney S, Greenhalgh T, Glover M, Blatch-Jones A. Models and applications for measuring the impact of health research: Update of a systematic review for the Health Technology Assessment Programme. Health Technol Assess. 2016;20:76.
7. Funding First. Exceptional Returns: The Economic Value of America's Investment in Medical Research. New York: Lasker Foundation; 2000.
8. Access Economics. Exceptional Returns: The Value of Investing in Health R&D in Australia. Canberra: Access Economics; 2003.
9. Access Economics. Exceptional Returns: The Value of Investing in Health R&D in Australia II. Canberra: Access Economics; 2008.
10. Deloitte Access Economics. Returns on NHMRC funded Research and Development. Canberra: Deloitte Access Economics; 2011.
11. Abate M, Vanni D, Pantalone A, Salini V. Cigarette smoking and musculoskeletal disorders. Muscles Ligaments Tendons J. 2013;3(2):63–9.
12. Schane RE, Ling PM, Glantz SA. Health effects of light and intermittent smoking: a review. Circulation. 2010;121(13):1518–22.
13. Viergever RF, Hendriks TCC. The 10 largest public and philanthropic funders of health research in the world: what they fund and how they distribute their funds. Health Res Policy Syst. 2016;14:12.
14. UK Clinical Research Collaboration. UK Health Research Analysis 2014. London: Medical Research Council; 2015.
15. Kiadaliri A, Woolf A, Englund M. Musculoskeletal disorders as underlying cause of death in 58 countries, 1986–2011: trend analysis of WHO mortality database. BMC Musculoskeletal Disorders. 2017;18:62.
16. Arthritis Research UK. Understanding Arthritis: A Parliamentary Guide to Musculoskeletal Health. Chesterfield: Arthritis Research UK; 2013. http://www.arthritisresearchuk.org/~/media/Files/Policy%20files/Policy%20pages%20files/Understanding%20Arthritis%20%20-%20Arthritis%20Research%20UK%20July%202013.ashx. Accessed Dec 2017.
17. Jordan KP, Joud A, Bergknut C, Croft P, Edwards JJ, Peat G, Petersson IF, Turkiewicz A, Wilkie R, Englund M. International comparisons of the consultation prevalence of musculoskeletal conditions using population-based healthcare data from England and Sweden. Ann Rheum Dis. 2013;73:212e8.
18. Woolf A, Pfleger B. Burden of major musculoskeletal conditions. Bulletin World Health Organ. 2003;81:9.
19. WHO. ICD-10 Version: 2016. http://www.who.int/classifications/icd10/browse/2016/en. Accessed March 2017.
20. Wellcome Trust. Putting NHS Research on the Map: An Analysis of Scientific Publications in England, 1990–97. 2001. https://wellcome.ac.uk/sites/default/files/wtd003193_0.pdf. Accessed March 2017.
21. Maddock I. The Seventh Royal Society Technology Lecture: Science, Technology and Industry. Proc R Soc Lond A Math Phys Sci. 1975;345(1642):295–326.
22. HM Treasury. GDP Deflators at Market Prices, and Money GDP: March 2015. 2015. https://www.gov.uk/government/statistics/gdp-deflators-at-market-prices-and-money-gdp-march-2015-budget-2015. Accessed March 2017.
23. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal 2013. 2013. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781. Accessed March 2017.
24. Murray C, Richards MA, Newton JN, Fenton KA, Anderson HR, Atkinson C, Bennett D, Bernabé E, Blencowe H, Bourne R, Braithwaite T, Brayne C, Bruce NG, Brugha TS, Burney P, Dherani M, Dolk H, Edmond K, Ezzati M, Flaxman AD, Fleming TD, Freedman G, Gunnell D, Hay RJ, Hutchings SJ, Ohno SL, Lozano R, Lyons RA, Marcenes W, Naghavi M, Newton CR, Pearce N, Pope D, Rushton L, Salomon JA, Shibuya K, Vos T, Wang H, Williams HC, Woolf AD, Lopez AD, Davis A. UK health performance: findings of the Global Burden of Disease Study 2010. Lancet. 2013;381:997–1020.
25. National Institute for Health and Care Excellence. Musculoskeletal Conditions Overview. https://pathways.nice.org.uk/pathways/musculoskeletal-conditions. Accessed March 2017.
26. National Collaborating Centre for Chronic Conditions. Osteoarthritis: National Clinical Guideline for Care and Management in Adults. London: Royal College of Physicians; 2008.
27. National Collaborating Centre for Chronic Conditions. Rheumatoid Arthritis: National Clinical Guideline for Management and Treatment in Adults. London: Royal College of Physicians; 2009.
28. Savigny P, Kuntze S, Watson P, Underwood M, Ritchie G, Cotterell M, Hill D, Browne N, Buchanan E, Coffey P, Dixon P, Drummond C, Flanagan M, Greenough C, Griffiths M, Halliday-Bell J, Hettinga D, Vogel S, Walsh D. Low Back Pain: Early Management of Persistent Non-Specific Low Back Pain. London: National Collaborating Centre for Primary Care and Royal College of General Practitioners; 2009.
29. National Institute for Health and Care Excellence. The Management of Hip Fracture in Adults. London: National Clinical Guideline Centre; 2011. https://www.nice.org.uk/guidance/cg124. Accessed March 2017.
30. Hospital Episode Statistics. http://content.digital.nhs.uk/hes. Accessed March 2017.
31. Prescription Cost Analysis Data. https://digital.nhs.uk/search?q=%22Prescription+Cost+Analysis+(PCA)+provides+details+of+the+number+of+items+and+the+net+ingredient+cost+%22&s=r. Accessed Dec 2017.
32. Prescribing Costs in Hospitals and the Community - National Level Time Series. https://data.gov.uk/dataset/hospital_prescribing_england. Accessed Dec 2017.
33. The Clinical Practice Research Datalink. https://www.cprd.com/intro.asp. Accessed March 2017.
34. Office for National Statistics: Population Estimates. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates. Accessed March 2017.
35. Curtis L, Burns A. PSSRU: Unit Costs of Health & Social Care 2015. http://www.pssru.ac.uk/project-pages/unit-costs/2015. Accessed March 2017.
36. Claxton K, Martin S, Soares M, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. Methods for the Estimation of the NICE Cost Effectiveness Threshold. Health Technol Assess. 2015;19:14.
37. Glover D, Henderson J. Quantifying Health Impacts of Government Policies. A How-to Guide to Quantifying the Health Impacts of Government Policies. Department of Health. 2010. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216003/dh_120108.pdf. Accessed March 2017.

38. Grant J, Cottrell R, Cluzeau F, Fawcett G. Evaluating "payback" on biomedical research from papers cited in clinical guidelines: applied bibliometric study. BMJ. 2000;320:1107–11.
39. Kinder AJ, Hassell AB, Brand J, Brownfield A, Grove M, Shadforth MF. The treatment of inflammatory arthritis with methotrexate in clinical practice: treatment duration and incidence of adverse drug reactions. Rheumatology. 2005;44(1):61–6.
40. Tosh JC, Wailoo AJ, Scott DL, Deighton CM. Cost-effectiveness of combination nonbiologic disease-modifying antirheumatic drug strategies in patients with early rheumatoid arthritis. J Rheumatol. 2011; 38(8):1593–600.
41. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, Hernandez-Alava M, Pasley S, Dickinson K, Scott D, Young A, Wailoo A. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. Health Technol Assess. 2016; 20(35):1–610.
42. Malottki K, Barton P, Tsourapas A, Uthman AO, Liu Z, Routh K, Connock M, Jobanputra P, Moore D, Fry-Smith A, Chen YF. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. Health Technol Assess. 2011;15(14):1–278.
43. Jackson R, Tosh J, Davis S, Wong R, Stevenson M, Stevens J. Golimumab for the Treatment of Rheumatoid Arthritis After Failure of Previous Disease-Modifying Antirheumatic Drugs: A Single Technology Appraisal. Sheffield: ScHARR; 2010.
44. Minton J, Tappenden P, Tosh P. Tocilizumab For the Treatment of Rheumatoid Arthritis: Report by the Decision Support Unit. School of Health and Related Research, University of Sheffield; 2011. http://docplayer.net/16963093-Tocilizumab-for-the-treatment-of-rheumatoid-arthritis.html. Accessed Dec 2017.
45. Shepherd J, Cooper K, Harris P, Picot J, Rose M. The clinical effectiveness and cost-effectiveness of abatacept, adalimumab, etanercept and tocilizumab for treating juvenile idiopathic arthritis: a systematic review and economic evaluation. NIHR Journals Library. 2016;20(34):234.
46. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, Myers L, Bruce I, Chalmers R, Bujkiewicz S, Lai M, Cooper N, Abrams K, Spiegelhalter D, Sutton A, Sculpher M, Woolacott N. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. Health Technol Assess. 2011;15(10):1–160.
47. Corbett M, Soares M, Jhuti G, Rice S, Spackman E, Sideris E, Moe-Byrne T, Fox D, Marzo-Ortega H, Kay L, Woolacott N, Palmer S. Tumour necrosis factor-α inhibitors for ankylosing spondylitis and non-radiographic axial spondyloarthritis: a systematic review and economic evaluation. Health Technol Assess. 2016;20(9):333.
48. Beard SM, von Scheele BG, Nuki G, Pearson IV. Cost-effectiveness of febuxostat in chronic gout. Eur J Health Econ. 2014;15(5):453–63.
49. National Joint Registry. Reports, Publications and Minutes. http://www.njrcentre.org.uk/njrcentre/Reports,PublicationsandMinutes/tabid/85/Default.aspx. Accessed March 2017.
50. Pennington M, Grieve R, Sekhon JS, Gregg P, Black N, van der Meulen JH. Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. BMJ. 2013;346:f1026.
51. De Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, Hutchison J, Grant A, Ccoyle D, Coyle K, Vale L. Systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip. Health Technol Assess. 2008;12(26):223.
52. National Collaborating Centre for Chronic Conditions (Great Britain), National Institute for Clinical Excellence (Great Britain), and Royal College of Physicians of London. Osteoarthritis: National Clinical Guidelines for Care and Management in Adults. London: Royal College of Physicians; 2008.
53. Wilson EC, Jayne DR, Dellow E, Fordham RJ. The cost-effectiveness of mycophenolate mofetil as firstline therapy in active lupus nephritis. Rheumatology. 2007;46(7):1096–101.
54. Stevenson M, Jones ML, De Nigris E, Brewer N, Davis S, Oakley J. A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis. Health Technol Assess. 2005;9(22):1–160.
55. National Institute for Health and Care Excellence. Denosumab for the Prevention of Osteoporotic Fractures in Postmenopausal Women. Technology Appraisal Guidance. TA204. 2010. https://www.nice.org.uk/guidance/ta204. Accessed March 2017.
56. National Institute for Health and Care Excellence. Alendronate, Etidronate, Risedronate, Raloxifene, Strontium Ranelate and Teriparatide for the Secondary Prevention of Osteoporotic Fragility Fractures in Postmenopausal Women. Technology Appraisal Guidance. TA161. 2008. https://www.nice.org.uk/guidance/ta161. Accessed March 2017.
57. National Institute for Health and Care Excellence. Low back pain in adults: early management. CG 88. 2009. [Withdrawn and replaced with NICE NG 59]
58. National Institute for Health and Care Excellence. Low Back Pain and Sciatica in Over 16s: Assessment and Management. NG 59. 2016. https://www.nice.org.uk/guidance/NG59. Accessed March 2017.
59. Clinical Standards Advisory Group. Back Pain. Report of a CSAG Committee on Back Pain. London: Her Majesty's Stationary Office; 1994.
60. UK Beam Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. BMJ. 2004;329:1377.
61. King's College London and Digital Science. The nature, scale and beneficiaries of research impact: an initial analysis of Research Excellence Framework (REF) 2014 impact case studies. Bristol: HEFCE; 2015. http://www.kcl.ac.uk/sspp/policy-institute/publications/Analysis-of-REF-impact.pdf. Accessed March 2017.
62. Wooding S, Hanney S, Buxton M, Grant J. The Returns from Arthritis Research. Volume 1: Approach, Analysis and Recommendations. MG251. Cambridge: RAND Europe; 2004. http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG251.pdf. Accessed March 2017.
63. Wooding S, Hanney S, Buxton MJ, Grant J. Payback arising from research funding: evaluation of the Arthritis Research Campaign. Rheumatology. 2005;44:1145–56.
64. Guthrie S, Grant J, Hanney S, Pollitt A. Investigating Time Lags and Attribution in the Translation of Cancer Research: A Case Study Approach. Cambridge: RAND Europe; 2014.
65. Hanney SR, Castle-Clarke S, Grant J, Guthrie S, Henshall C, Mestre-Ferrandiz J, Pistollato M, Pollitt A, Sussex J, Wooding S. How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice. Health Res Policy Syst. 2015;13:1.
66. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med. 2011;104: 510–20.
67. Adams J, Gurney K. The Implications of International Research Collaboration for UK Universities. London: Digital Science; 2016. https://www.digital-science.com/resources/digital-research-reports/digital-research-report-the-implications-of-international-research-collaboration-for-uk-universities/. Accessed March 2017.
68. Department of Business, Innovation and Skills. International Comparative Performance of the UK Research Base – 2013: Appendix F – supplementary data. 2013. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263763/international-comparative-performance-of-the-UK-research-base-2013-appendix-f-supplementary-data.pdf. Accessed March 2017.
69. O'Brien S, Bennett D, Doran E, Beverland DE. An outcome study on average length of stay following total hip and knee replacement. J Orthop Nurs. 2002;6(3):161–9.
70. Iyengar KP, Nadkarni JB, Ivanovic N, Mahale A. Targeted early rehabilitation at home after total hip and knee joint replacement: does it work? Disabil Rehabil. 2007;29(6):495–502.
71. Mertes SC, Raut S, Khanduja V. Integrated care pathways in lower-limb arthroplasty: are they effective in reducing length of hospital stay? Int Orthod. 2013;37(6):1157–63.

APPENDIX H: Co-author declarations

Given the nature of health research, several of the works were chapters in much larger publications. Some of these contained clinical, statistical and other components in addition to the health economic analyses. Some are multi-authored to reflect various contributions from those in administrative, management and research roles to overall study conduct (e.g. P4 - 15 authors, P2 - 8 authors). This is common in health research including modelling studies and randomised controlled trials, where research design and execution is complex and multidisciplinary.

To gather sufficient evidence of candidate contribution but increase efficiency in acquiring appropriate declarations, a system was developed to gather statements from those most closely involved in the health economic component. In some papers I appear as a part of a longer list of co-authors who contributed to the overall project, but my involvement is that of either a 'lead', 'the principal analyst' of the health economic component and/or first/corresponding author as outlined in Table 1.

As this role required working closely with the Principal Investigator (PI) of funded studies and Lead/Corresponding authors (where not PI), sometimes under supervision by a health economist colleague, these are the most relevant co-authors to confirm my contribution to a particular paper. To create a consistent approach, for each published work a declaration of candidate contribution was therefore signed by:

1. Principal Investigator of each of the related studies from which the work developed
2. First authors and corresponding authors of the published works, where these were different from the Principal Investigator (and where I did not fulfil these roles)
3. Senior colleagues leading supervision of economic components of related studies or senior institutional colleagues (where I did not fulfil this role)

If this system did not produce a minimum of two declarations for any of the works, a declaration was sought from another colleague closely involved in the work. Therefore, the papers concerned with AAA screening (P1, P2, P3) all have an extra declaration from an additional author. The roles fulfilled by co-authors signing declarations is summarised below in Table 7 and Table 8.

Table 7: Co-author declarations summary

Paper	**Co-author declarations**					**Number of declarations**
	Principal Investigator	First author	Corresponding author	Senior Institutional (BUL) author	Additional author	
P1 NHS NAAASP	ST	MG	MG	MB	MS	3
P2 AAA surveillance intervals	ST	ST	ST	MB	MS	3
P3 Lessons from AAA modelling	ST	MG	MG	N/A	MS	2
P4 TOMADO	TQ	LS	LS	JFR		3
P5 HTA impact review	JR	JR	JR	SH		2
P6 Est returns to cancer research	JG	MG	JG	MB		2
P7 Est returns to MSK research	JG	MG	JG	MB		2

ST: Prof Simon Thompson; TQ: Dr Timothy Quinnell, LS: Prof Linda Sharples; JR: Prof James Raftery; MG: Matthew Glover; JG: Prof Jonathan Grant; MB: Prof Martin Buxton; JFR: Prof Julia Fox-Rushby; SH: Prof Stephen Hanney; MS: Dr Michael Sweeting

Table 8: List of co-authors signing declarations

Paper	**Co-author declarations**
P1 NHS NAAASP	Professor Simon Thompson (PI), Professor Martin Buxton (SBUL), Dr Michael Sweeting (AA)
P2 AAA surveillance intervals	Professor Simon Thompson (PI), Professor Martin Buxton (SBUL), Dr Michael Sweeting (AA)
P3 Lessons from AAA modelling	Professor Simon Thompson (PI), Dr Michael Sweeting (AA)
P4 TOMADO	Dr Timothy Quinnell (PI), Professor Linda Sharples (FA), Professor Julia Fox-Rushby (SBUL)
P5 HTA impact review	Professor James Raftery (PI), Professor Stephen Hanney (SBUL)
P6 Est returns to cancer research	Professor Jonathan Grant (PI), Professor Martin Buxton (SBUL)
P7 Est returns to MSK research	Professor Jonathan Grant (PI), Professor Martin Buxton (SBUL)

PI: Principal Investigator; SBUL: Senior Brunel University London colleague; AA: Additional Co-author; FA: First author

Appendix H1 – Paper 1 co-author declarations

Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England. Br J Surg. 2014;101(8):976-982. doi:10.1002/BJS.9528

Co-author declarations:

- Professor Simon Thompson
- Professor Martin Buxton
- Dr Michael Sweeting

Co-author declaration of candidate contribution
Paper number: 1
Title: Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England.
Published in: British Journal of Surgery
Candidate contribution: Led update of model including input parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), performed cost-effectiveness analyses. Contributed to conception of research question. Led preparation and submitted manuscript. Responsible for manuscript revisions, including additional cost-effectiveness analyses. First and corresponding author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work. Signed: Simon G. Thompson ..... Print name: Professor Simon G Thompson ..... Role in publication: One of two senior authors; recognised the importance of the research question; helped identify relevant data sources; provided funding for Cambridge authors. Job title: Director of Research in Biostatistics, University of Cambridge. .....

Co-author declaration of candidate contribution
Paper number: 1
Title: Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England.
Published in: British Journal of Surgery
Candidate contribution: Led update of model including input parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), performed cost-effectiveness analyses. Contributed to conception of research question. Led preparation and submitted manuscript. Responsible for manuscript revisions, including additional cost-effectiveness analyses. First and corresponding author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: ..... Signed: M. J. Buxton ..... Print name: Martin J Buxton ..... Role in publication: Conceived and set-up project and oversaw and advised on cost-effectiveness analysis. Contributed to structure and drafting of manuscript. .....

Co-author declaration of candidate contribution

Paper number: 1

Title: Cost-effectiveness of the National Health Service abdominal aortic aneurysm screening programme in England.

Published in: British Journal of Surgery

Candidate contribution: Led update of model including input parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), performed cost-effectiveness analyses. Contributed to conception of research question. Led preparation and submitted manuscript. Responsible for manuscript revisions, including additional cost-effectiveness analyses. First and corresponding author.

I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work:

……..

I can confirm that Mr Glover's contributions are accurately described.

……..

……..

……..

Signed:

……..

Print name:
Dr Michael Sweeting

……..

Role in publication:

During this work I was a Senior Research Associate at University of Cambridge. The work submitted was a cost-effectiveness analysis conducted by Mr Glover and was part of a wider HTA funded grant involving a systematic review and meta-analysis of growth and rupture rates of abdominal aortic aneurysm. My contribution to this work was to provide input data and statistical consultancy for certain aspects of the modelling, consult and review model outputs and to critically review and comment on manuscript drafts.

Appendix H2 – Paper 2 co-author declarations

Thompson SG, Brown LC, Sweeting MJ, Bown MJ, Kim LG, **Glover MJ,** Buxton MJ, Powell JT and the RESCAN collaborators. Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness. Health Technol Assess. 2013;17(41):1-118. doi:10.3310/HTA17410 **(Chapters 7 and 8 of larger monograph)**

Co-author declarations:

- Professor Simon Thompson
- Professor Martin Buxton
- Dr Michael Sweeting

Co-author declaration of candidate contribution
Paper number: 2
Title: Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness.
Published in: Health Technology Assessment
Candidate contribution: Led update of model parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), developed framework and performed structural amendments to operationalised model. Conducted cost-effectiveness analyses, drafted chapters 7 and 8 and contributed to revisions, including additional cost-effectiveness analyses. Contributed to critical revision of other relevant sections of monograph.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work. Signed: Simon G. Thompson ..... Print name: Professor Simon G Thompson ..... Role in publication: First and corresponding author of the whole monograph (of which chapters 7 and 8 are a part); lead investigator obtaining competitive funding for project; oversight of all components of the project. Job title: Director of Research in Biostatistics, University of Cambridge. .....

Co-author declaration of candidate contribution
Paper number: 2
Title: Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness.
Published in: Health Technology Assessment
Candidate contribution: Led update of model parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), developed framework and performed structural amendments to operationalised model. Conducted cost-effectiveness analyses, drafted chapter 7 and 8 and contributed to revisions, including additional cost-effectiveness analyses. Contributed to critical revision of other relevant sections of monograph.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: ..... Signed: M. J. Buxton ..... Print name: Martin J Buxton ..... Role in publication: Involved in conception of research and co-applicant on grant. Supervised work on model and contributed to drafting of manuscript. .....

Co-author declaration of candidate contribution
Paper number: 2
Title: Systematic review and meta-analysis of the growth and rupture rates of small abdominal aortic aneurysms: implications for surveillance intervals and their cost-effectiveness.
Published in: Health Technology Assessment
Candidate contribution: Led update of model parameters (including data identification/acquisition of key screening programme data, costs, health related quality of life and other health economic inputs), developed framework and performed structural amendments to operationalised model. Conducted cost-effectiveness analyses, drafted chapter 7 and 8 and contributed to revisions, including additional cost-effectiveness analyses. Contributed to critical revision of other relevant sections of monograph.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: …….. I can confirm that Mr Glover's contributions are accurately described. Matthew took the lead role in conducting cost-effectiveness analyses for this HTA monograph. Signed:  … Print name: Dr Michael Sweeting …….. Role in publication: During this work I was a Senior Research Associate at University of Cambridge. The work submitted is an HTA monograph from a HTA funded grant involving a systematic review and meta-analysis of growth and rupture rates of abdominal aortic aneurysm (AAA) and cost-effectiveness analysis of AAA screening and surveillance policies. My contribution to this work involved assimilating and cleaning of individual patient data from observational studies, conducting meta-analyses, providing input data on growth and rupture rates for cost-effectiveness models, summarising and dissemination of results through publications, and providing overall statistical consultancy throughout the project. I drafted Chapters 5 and 6 of the HTA monograph and critically reviewed and commented on all chapters of the report. I worked closely with Matthew throughout the project.

Appendix H3 – Paper 3 co-author declarations

Glover MJ, Jones E, Masconi KL, et al. Discrete Event Simulation for Decision Modeling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening. Med Decis Mak. 2018;38(4):439-451. doi:10.1177/0272989X17753380

Co-author declarations:

- Professor Simon Thompson
- Dr Michael Sweeting

Co-author declaration of candidate contribution
Paper number: 3
Title: Discrete Event Simulation for Decision Modelling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening.
Published in: Medical Decision Making
Candidate contribution: Co-investigator on National Institute for Health Research Health Technology Assessment grant, involved in conception of research question. Contributed health economics input to conceptual model development and led update of costs and utility data (including identification and acquisition of data). Conceived, led preparation and submitted manuscript and responsible for revisions. Listed as first and corresponding author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work. Signed: Simon G. Thompson ..... Print name: Professor Simon G Thompson ..... Role in publication: Senior author; lead investigator obtaining funding for overall project (Screening women for AAA), of which this was an initial component; direct supervision of Cambridge co-authors. Job title: Director of Research in Biostatistics, University of Cambridge. .....

Co-author declaration of candidate contribution

Paper number: 3

Title: Discrete Event Simulation for Decision Modelling in Health Care: Lessons from Abdominal Aortic Aneurysm Screening.

Published in: Medical Decision Making

Candidate contribution: Co-investigator on National Institute for Health Research Health Technology Assessment grant, involved in conception of research question. Contributed health economics input to conceptual model development and led update of costs and utility data (including identification and acquisition of data). Conceived, led preparation and submitted manuscript and responsible for revisions. Listed as first and corresponding author.

I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work:

……..

I can confirm that Mr Glover's contributions are accurately described.

……..

……..

Signed:

……..

Print name:
Dr Michael Sweeting

……..

Role in publication:

… During this work I was a Senior Research Associate at University of Cambridge. I was a co-investigator on this NIHR HTA grant and was involved in the conception of the research question, model development, sourcing and acquisition of data, reviewing statistical programming and line management of a research associate. My contribution to this manuscript involved conception, provision of key data and model results, and critical review and revision of the text……..

Appendix H4 – Paper 4 co-author declarations

Sharples L, **Glover M**, Clutterbuck-James A, Bennett M, Jordan J, Chadwick R, Pittman M, East C, Cameron M, Davies M, Oscroft N. Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airway pressure. Health technology assessment (Winchester, England). 2014 Oct;18(67):1. **(Chapter 4 of larger monograph)**

Co-author declarations:

- Dr Timothy Quinnell
- Professor Linda Sharples
- Professor Julia Fox-Rushby

Co-author declaration of candidate contribution

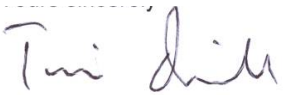
Paper number: 4

Title: Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airway pressure.

Published in: Health Technology Assessment

Candidate contribution: Contributed to systematic review of effectiveness, including data extraction and input to statistical analyses performed for health economic model inputs. Led structured review of literature to inform model amendments, updated model parameters and conducted cost-effectiveness analyses. Drafted Chapter 4 and made revisions, including performing additional cost-effectiveness analyses. Contributed to critical revision of other sections of monograph and helped supervise within-trial economic analysis.

I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work:

Signed: 

……..

Print name:
Timothy G. Quinnell

……..

Role in publication:
I was Chief Investigator for TOMADO. At the time of the study and to this date I have been based at the study site (Royal Papworth Hospital NHS Foundation Trust) as a consultant respiratory and sleep disorders physician. I led the NIHR grant application for TOMADO. I also led on study design, execution and results interpretation/dissemination. I was lead author for the journal article (Thorax) reporting the results of the study. I was senior author for the NIHR report. I confirm that the candidate Matthew Glover was a key team member. He contributed as detailed above.

Co-author declaration of candidate contribution
Paper number: 4
Title: Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airway pressure.
Published in: Health Technology Assessment
Candidate contribution: Contributed to systematic review of effectiveness, including data extraction and input to statistical analyses performed for health economic model inputs. Led structured review of literature to inform model amendments, updated model parameters and conducted cost-effectiveness analyses. Drafted Chapter 4 and made revisions, including performing additional cost-effectiveness analyses. Contributed to critical revision of other sections of monograph and helped supervise within-trial economic analysis.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: Signed: L Sharples. …….. Print name: Professor Linda Sharples …….. Role in publication: During conduct of this trial I was employed as a Programme leader at the MRC Biostatistics Unit in Cambridge, which was part-funded by Papworth Hospital. By the time it was published I had moved to become Professor of Statistics at the University of Leeds Clinical Trials Research Unit. I was a co-applicant on the study grant application and had overall responsibility for the statistical design and conduct of the study (PI for statistical and research methodology). I was responsible for the statistical design and analysis and drafted large sections of the monograph. With clinical and health economic collaborators I oversaw production of the final publication.

Co-author declaration of candidate contribution

Paper number: 4

Title: Clinical effectiveness and cost-effectiveness results from the randomised controlled Trial of Oral Mandibular Advancement Devices for Obstructive sleep apnoea-hypopnoea (TOMADO) and long-term economic analysis of oral devices and continuous positive airway pressure.

Published in: Health Technology Assessment

Candidate contribution: Contributed to systematic review of effectiveness, including data extraction and input to statistical analyses performed for health economic model inputs. Led structured review of literature to inform model amendments and updated model parameters and conducted cost-effectiveness analyses. Drafted chapter 4 and made revisions, including performing additional cost-effectiveness analyses. Contributed to critical revision of other sections of monograph and helped supervise within-trial economic analysis presented in Chapter 2.

I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work:

.....

.....

Signed:
J.A. Fox-Rushby
.....

Print name:
Julia Fox-Rushby, Professor of Health Economics, Kings College London
.....

Role in publication:

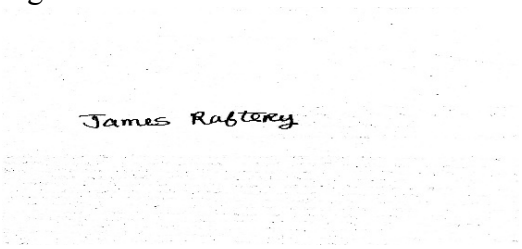
I was Prof Health Economics and Director of HERG at the time Matt completed this work. I had overall responsibility for the health economics. The original grant funding had been won by Prof Buxton and data collection set up by Dr Gethin Griffiths. I took over responsibility for grant completion for economics on Prof Buxton's retirement and Matthew took on the economic modelling, and provided additional day to day support for the other economics researcher completing his first trial analysis. Matt brought new ideas to the analysis of quality of life trial data for trial analysis. I reviewed and discussed analysis plans, model development plans, results and contributed to parts of the first draft for the trial analysis. Matt completed the first drafts for all aspects of the economic modelling and its supporting literature review. He also jointly conducted the systematic review of effectiveness evidence.

Appendix H5 – Paper 5 co-author declarations

Raftery J, Hanney S, Greenhalgh T, **Glover M**, Blatch-Jones A. Models and applications for measuring the impact of health research: Update of a systematic review for the health technology assessment programme. Health Technol Assess. 2016;20(76):1-282. doi:10.3310/hta20760 **(Chapter 5 of larger monograph)**

Co-author declarations:

- Professor James Raftery
- Professor Stephen Hanney

Co-author declaration of candidate contribution
Paper number: 5
Title: Models and applications for measuring the impact of health research: update of a systematic review for the Health Technology Assessment programme.
Published in: Health Technology Assessment
Candidate contribution: Contributed to systematic review in Chapter 3 (including data extraction), led systematic review of economic impact assessment literature (project lead on Chapter 5), drafted Chapter 5 and revisions. Contributed to critical revision of other sections of monograph.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: …….. Matthew Glover was an excellent colleague who performed the above tasks to a very high standard …….. …….. Signed:  …….. Print name: James Raftery 20 February 2022 …….. Role in publication: Lead investigator, grant holder and other author ……..

Co-author declaration of candidate contribution
Paper number: 5
Title: Models and applications for measuring the impact of health research: update of a systematic review for the Health Technology Assessment programme.
Published in: Health Technology Assessment
Candidate contribution: Contributed to systematic review in Chapter 3 (including data extraction), led systematic review of economic impact assessment literature (project lead on Chapter 5), drafted Chapter 5 and revisions. Contributed to critical revision of other sections of monograph.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: This accurately reflects Matthew's role in this HTA review in which the team worked together overall, but took a lead on different chapters. Matthew was clearly the lead on chapter 5 which had its own systematic review Signed: Stephen R Hanney Print name: STEPHEN R. HANNEY Role in publication: Led on writing chapter 3, contributed to other sections

Appendix H6 – Paper 6 co-author declarations

Glover M, Buxton M, Guthrie S, Hanney S, Pollitt A, Grant J. Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. BMC Med 2014 121. 2014;12(1):1-21. doi:10.1186/1741-7015-12-99

Co-author declarations:

- Professor Jonathan Grant
- Professor Martin Buxton

Co-author declaration of candidate contribution
Paper number: 6
Title: Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes.
Published in: BMC Medicine
Candidate contribution: Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and revisions, including performing additional modelling. Contributed critical revision to other sections of manuscript. Listed as first author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: …….. …….. Signed:  Print name: Dr Jonathan Grant Role in publication: I was the co-PI/lead investigator of this grant and corresponding author for this paper when, at the time, professor of public policy at King's College London. I can confirm that Matthew was a critical member of the team making the contributions outlined above and it is entirely appropriate that he includes this paper as part of a PhD submission.

Co-author declaration of candidate contribution
Paper number: 6
Title: Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes.
Published in: BMC Medicine
Candidate contribution: Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and revisions, including additional performing modelling. Contributed critical revision to other sections of manuscript. Listed as first author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: ..... Signed: ..... Print name: Martin J Buxton ..... Role in publication: Co-leader in conception of project and obtaining funding and support, Steered and supervised work on economic returns and contributed to structuring and drafting manuscript. .....

Appendix H7 – Paper 7 co-author declarations

Glover M, Montague E, Pollitt A, et al. Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes. Heal Res Policy Syst 2018 161. 2018;16(1):1-24. doi:10.1186/S12961-017-0276-7

Co-author declarations:

- Professor Jonathan Grant
- Professor Martin Buxton

Co-author declaration of candidate contribution
Paper number: 7
Title: Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes.
Published in: Health Research Policy and Systems
Candidate contribution: Co-investigator on Wellcome Trust led grant, involved in conception of research question. Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and performed revisions, including additional modelling. Contributed critical revision to other sections of manuscript. Listed as first author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: …….. …….. …….. …….. Signed:  Print name: Dr Jonathan Grant Role in publication: I was the co-PI/lead investigator of this grant and corresponding author for this paper when, at the time, professor of public policy at King's College London. I can confirm that Matthew was a critical member of the team making the contributions outlined above and it is entirely appropriate that he includes this paper as part of a PhD submission.

Co-author declaration of candidate contribution
Paper number: 7
Title: Estimating the returns to United Kingdom publicly funded musculoskeletal disease research in terms of net value of improved health outcomes.
Published in: Health Research Policy and Systems
Candidate contribution: Co-investigator on Wellcome Trust led grant, involved in conception of research question. Led prioritisation of interventions, identification and collation of data on net health gains, including liaison with expert contributors. Developed return on investment model. Drafted sections relating to economic model methods and results and performed revisions, including additional modelling. Contributed critical revision to other sections of manuscript. Listed as first author.
I hereby confirm that the above statement is a fair reflection of the candidate contribution in producing the published work: ..... Signed: M. J. Buxton ..... Print name: Martin J Buxton ..... Role in publication: Co-leader in conception of project and obtaining funding and support, Steered and supervised work on economic returns and contributed to structuring and drafting manuscript. .....