Strategy Selection and Comparison in the Cost-Effectiveness Analysis of Cancer Screening

Yi-Shu Lin, MSc, MBA

A dissertation submitted for the Degree of Doctor of Philosophy

School of Medicine

Trinity College Dublin

Under the supervision of Dr. James O'Mahony

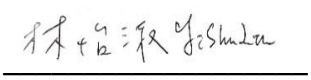
2023

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).



Yi-Shu Lin

Summary

Cost-effectiveness analysis (CEA) is well-established in application to cancer screening programmes. CEAs of screening differ from analyses of therapeutic interventions as the choice of strategies for comparison is not limited simply by the different interventions to compare but also the intensity at which they are provided. This very much depends on modelling choices made by the analysts conducting the simulation study. Existing checklists and guidelines for the application of CEA fail to fully address important issues regarding strategy choice which are relevant for adequate incremental cost-effectiveness ratio (ICER) estimation. Applied CEA models are an important tool for assessing health and economic effects of healthcare interventions but are not best suited for illustrating methodological issues. The objectives of this thesis are to review the issue of strategy omission in the CEAs of cancer screening, to provide a simple, open-source model for the simulation of disease screening cost-effectiveness for teaching and research purposes, and to demonstrate how the omission of strategy can arise by observing the impacts of characteristics of screening schedules and parameter values on the shape and composition of the efficient frontier.

This thesis comprises three studies. The first identified the issue of strategy omission existing in the CEAs of colorectal cancer (CRC) screening. The quality assessment tool developed by this thesis identified methodological concerns around the ICER generation due to inappropriate comparisons and inadequate comparators. This study revealed that many CEAs did not consider different types, i.e., stool-based and image-based, of screening modalities available for CRC screening. Many CEAs did not consider consistent screening schedules for the same type of screening modalities. In addition, this study found around half of the CEAs reviewed reported inappropriate ICERs for decision-making. In most cases, ICERs are confused with average cost-effectiveness ratios (ACERs), which are based on the comparison to the no-screening scenario. Another common inappropriate way of reporting is a cross-tabulated table that listed all the cost-effectiveness ratios (CERs), comparing all strategies to every other strategy.

The second study described a microsimulation model expressly designed for methods demonstration in CEAs of cancer screening. This simplified model has rapid simulation time, allowing the simulation of a large range of screening strategies. This framework is highly flexible in adjusting the intensities of screening strategies. This model can be used

to demonstrate the complexity of screening strategies and their cost-effectiveness. The scenario analysis of this study showed how to conduct a simple face validation via the observation of the impact of changes in parameter values on the cost and effectiveness estimates. This study also showed the relevance of considering both gross cost and effects estimates and relative outcomes net to a no screening scenario. This distinction between gross and relative outcomes is helpful in providing an intuitive understanding of how cost-effectiveness estimates vary with changes in parameter values.

The third study used the pedagogical model proposed by the second study to demonstrate examples of strategy omission. The third study presented the trajectory of strategies with common characteristics on the cost-effectiveness plane with a variation in screening start ages, stop ages, and screening intervals. This study explained the policy-relevant portion of the efficient frontier which is comprised of strategies with ICERs close to the cost-effectiveness threshold and therefore should be of greatest interest to policy makers. Moreover, this study examined the convexity of the efficient frontier and whether this varies with changes in parameter values. In contrast to the position of efficient frontiers which gives a general understanding of the cost-effectiveness of screening relative to the no-screening, the change in the convexity implies the inconsistent change in ICERs of screening strategies with different intensities. This study provides guidance on how to avoid the omission of relevant strategies when conducting CEAs of cancer screening and suggests using an iterative method when choosing strategies to simulate. This approach uses the findings from the initial simulation to inform what other strategies might be worth adding to the simulation in order to include the most relevant strategies. The findings of this study give insights into how to expand the choice of strategies closer to the efficient frontier, improving the chance of strategies being cost-effective.

Take as a whole, this thesis describes the issue of strategy omission in cancer screening CEAs in detail. It examines the issue within the published literature and within an abstracted model. Using insights from both the review and a simplified model it provides suggestions on how best to avoid the omission of relevant screening strategies. Future research can extend this thesis to establish more detailed and realistic guidance for CEAs of screening.

Acknowledgements

From the bottom of my heart, I would like to express my sincere gratitude to my supervisor Dr. James O'Mahony who made this work possible. He continuously provided encouragement and was always willing and enthusiastic to facilitate in any way he could during my doctoral journey as well as my professional advancement. I am and I always will be proud of being his first PhD student.

Thanks to Dr. Joost van Rosmalen and Dr Stefan Lhachimi for their advice on the model development.

I also would like to say special thanks to Rajani, Yuri, Kathleen, and other PhD scholars in the centre. You have been my biggest cheerleaders, always listening to me, guiding me and making me believe in myself throughout the academic years. I would also like to thank all my Taiwanese friends in Ireland. Without your company, I might have not been able to complete my study abroad. Thank you to Zlatko for encouraging me when I am in the dark and supporting me whenever I need.

Finally, thank you to my dear family who has always been there for me. I am deeply touched by their own way of caring and love no matter where I am in the world. I am grateful in more ways than I can express for their unconditional support in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACER	average cost-effectiveness ratio
BE	barium enema
CADTH	Canadian Agency for Drugs and Technologies in Health
CEA	cost-effectiveness analysis
CEAC	cost-effectiveness acceptability curve
CER	cost-effectiveness ratio
CHEC-list	Consensus on Health Economic Criteria checklist
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CISNET	Cancer Intervention and Surveillance modelling NETwork
CMA	cost-minimisation analysis
CMOST	Colon Modelling Open Simulation Tool
CRAN	Comprehensive R Archive Network
CRC	colorectal cancer
CRC-SPIN	Colorectal Cancer Simulated Population model for Incidence and Natural history
CTC	computed tomography colonography
CUA	cost-utility analysis
DARTH	Decision Analysis in R for Technologies in Health
DES	discrete event simulation
Drummond checklist	Drummond Ten Point checklist
EVPPI	the expected value of partially perfect information
fDNA	faecal DNA
FIT	faecal immunochemical test
FOBT	faecal occult blood test
FHCRC	Fred Hutchinson Cancer Research Centre
GUI	graphical user interface
HIQA	Health Information and Quality Authority
HPV	human papillomavirus
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio
IPECAD	International Pharmaco-Economic Collaboration on Alzheimer's Disease
ISPOR	Professional Society for Health Economics and Outcomes Research

IVI	Innovation and Value Initiative
LMICs	low and middle income countries
LYs	life-years
MCDA	multi-criteria decision analyses
MISCAN	MIcrosimulation SCreening ANalysis
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMB	net monetary benefit
NSCLC	non-small cell lung cancer
OSVP	Open-Source Value Project
OWSA	one-way sensitivity analysis
PSA	probabilistic sensitivity analysis
PBAC	Pharmaceutical Benefits Advisory Committee
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	international PROSPEctive Register Of systematic reviews
QALYs	quality-adjusted life-years
QoL	quality-of-life
QHES checklist	Quality of Health Economic Studies checklist
RA	rheumatoid arthritis
ScHARR	the University of Sheffield School of Health and Related Research
SimCRC	Simulation Model of Colorectal Cancer
SMDM	Society for Medical Decision Making
VBA	Visual Basic for Applications
WHO	World Health Organization

1. INTRODUCTION

Cost-effectiveness analysis (CEA) seeks to inform a wide range of different and unavoidable decisions in healthcare due to limited resources (Drummond et al., 2015). To maximize health benefits from finite resources, the decision-maker needs to compare the opportunity costs among the health programmes, including new interventions and existing technologies. CEA is commonly used as a research tool to decide if decision-makers should accept a specific health intervention. CEA is now an accepted part of healthcare resource allocation in many countries as reflected by the multitude of analysis guidelines published around the world (the Professional Society for Health Economics and Outcomes Research [ISPOR], 2022).

The term cost-utility analysis (CUA) is sometimes uses to represent the special case of a CEA in which quality-adjusted life-years (QALYs) rather than life-years (LYs) have been used as a measure of effectiveness. This thesis will follow the convention used by others which considers CUA and CEA to be synonymous and uses the latter term.

Cancer screening is a secondary prevention measure that can reduce the burden of cancer by intervening at an early stage of disease development (Eddy, 1986). Cancer screening has proved that it can effectively improve cancer survival in certain applications (Arbyn et al., 2009; Lăără et al., 1987). In 1968, Wilson and Jungner established the principles of screening for the World Health Organization (WHO), which stated that screening should have scientific evidence of effectiveness, especially which treatment should be available (Wilson & Jungner, 1968). An early example of a nationwide, organised screening programme is Sweden's national cervical cancer screening programme that began in 1967 (Basu et al., 2018). So far, most European countries have initiated a variety of organised population-based screening programmes (Basu et al., 2018; Ponti et al., 2017). In 2008, Andermann et al. updated the WHO criteria for screening to include the consideration that the overall benefits should outweigh the harm, including an explicit consideration of screening costs relative to the alternative uses of resources elsewhere in the health system (Andermann et al., 2008).

CEA has been applied in cancer screening to evaluate its potential benefits and harms for decades. An early example is Dickinson (1972) that estimated $14 per LY gained after considering the cost of cytology examinations, the cost-saving from the early treatment, and the income tax that the Federal Government that can receive by extending LYs. Another example is Schweitzer (1974) that suggested one-time cervical cancer screening is cost-effective from the social perspective with an applied decision model. The book published by Eddy (1980) may be the most systematic and comprehensive early work on CEA in cancer screening. Eddy developed a general theory of screening and fitted available data to his proposed model with examples of cervical, breast, and colorectal cancer (CRC) screening (Eddy, 1980; Eddy, 1981).

One important difference between cancer screening and other interventions is the wide range of potential intervention intensities in the case of screening. Screening intervention can be implemented with a wide variety of screening age ranges and alternative screening intervals. Screening can also involve various testing modalities for both primary screening and triage testing. Screening policies sometimes vary between different risk populations. This flexibility makes the CEAs of screening very different from other CEAs designed for therapeutic interventions. The choice of strategy very much depends on what the modellers choose to include as the range of relevant strategies is not simply defined by the range of existing technologies such as current therapies and possibly their combinations.

CEAs of screening typically rely on simulation modelling rather than being based directly on clinical trials (Knudsen et al., 2007). The potential to vary aspects of screening intensity results in a large number of potential strategies which are infeasible to assess in trials. Trials can be expensive and require a long time to both implement and observe the performance of screening interventions (Knudsen et al., 2007). With CEA models, more screening strategies can be considered and estimates can be provided earlier compared to only relying on trial evidence (Mandelblatt et al., 1996, pp. 159-160). Accordingly, CEA models are often used to guide screening policy (Saslow et al., 2002; Smith et al., 2003).

A critical methodological problem in CEAs of screening has been reported, where relevant strategies are insufficiently explored, leading to a biased choose of optimal strategy (Fabbro et al., 2022; O'Mahony, 2014, 2015, 2019). When a limited range of strategy choices is considered corresponding with a narrow study question, the findings merely provide "cost-effectiveness ratio" estimates based on comparisons to strategies that may not be truly cost-effective. This misalignment with the purpose of CEA undermines the utility of such estimates. An essential contribution of economic evaluation is to mitigate the risk of overlooking important relevant alternatives that should be taken into consideration (Drummond et al., 2015, p. 45).

To illustrate the methodological challenge concerning strategy selection in cancer screening CEAs, this thesis reviews the existing literature on CEA methods applied to cancer screening. The following introduction is organized into three parts, each summarizing relevant literature pertaining to distinct topics. Part 1 focuses on the materials discussing the methods for choosing relevant screening strategies for simulation with an analysis. This includes suggestions from checklists, textbooks, and government documents for the appraisal of economic evaluation, along with the finding of recent systematic reviews of cancer screening. Part 2 presents aspects of the technical methods of CEA, including different model types in cancer screening, examples of open-source applied models and tutorials, and modelling tools. The reason for reviewing the simulation techniques is that some CEAs may intentionally simulate a limited range of strategies due to technical difficulties (Drummond et al., 2015, pp. 23-24). Part 3 reviews the common features among examples of CEAs simulating a wide array of screening strategies.

Part 1 CEA methods for the appraisal of cancer screening

Cost-effectiveness evidence of screening supports healthcare decision-making. The formulation of this evidence rests on a foundation of established methods for economic evaluation. It is important to adhere to these methods to maintain the quality of screening CEA estimates. If analysts diverge from methods guidance, then they may produce estimates which guide policy makers to inefficient policies, which in turn compromises population health. This section will introduce how to identify the relevant strategies within the CEAs of cancer screening, by explaining some fundamental methods of CEA,

listing relevant guidelines and checklists, describing possible variations for defining screening strategies, and identifying the methodological concerns raised in the recent existing systematic reviews of cancer screening CEAs.

Decision-making with CEA

In the theory of economics, optimal decision-making is made based on the estimation of the opportunity cost of any particular choice. "The opportunity cost of using an input to implement a policy is its value in its best alternative use" (Boardman et al., 2014, p. 31). In health economics, the corresponding tool for the concept of opportunity costs is the cost-effectiveness threshold. This represents the maximum cost per QALY that is considered justified. An intervention's incremental cost-effectiveness ratio (ICER) is compared to the cost-effectiveness threshold to decide whether to adopt the new intervention. The standard decision rule is to adopt the intervention with the highest ICER within the threshold. The ICER itself represents the net resource costs (numerator) divided by the net health effectiveness (denominator), which "net" changes are measured from the comparison to some baseline scenario (Torrance et al., 1972; Weinstein, 1990).

ICERs are only estimated for those strategies that fall within the efficient set of interventions within the cost-effectiveness plane, also known as the "efficient frontier" (Stinnett et al., 1996). The efficient frontier comprises those interventions with costs and effects estimates that are not subject to simple or extended dominance. A strategy of interest is subject to simple (strong) dominance if there exists one or more alternative interventions that are more effective and less costly than the strategy of interest (Drummond et al., 2015, pp. 98-101). A strategy of interest is subject to extended dominance if a hypothetical combination of two alternative treatments provided to the patient population result in greater effects and less costs than the strategy of interest (Siegel et al., 1996, p. 286).

ICERs are generated by calculating the ratio of the incremental cost difference between the strategy of interest and its comparator strategy to the increment difference in effects between the two strategies. The appropriate comparator strategy is the next most effective strategy on the efficient frontier relative to the strategy of interest (Siegel et al., 1996).

The concept of the ICER is often confused with the average cost-effectiveness ratio (ACER) (Drummond & Sculpher, 2005). The interpretation of ACER can be ambiguous and controversial in the context of screening. First, ACER generally refers to the ratio of the total costs divided by the total effectiveness for each intervention, rather than the incremental comparison among efficient interventions as with the ICER (Drummond & Sculpher, 2005). This ACER is on a basis of the comparison to a hypothetical scenario with absolute zero-cost and zero-effectiveness, such as no treatment. The interpretation within screening is typically that ACER can be considered relative to a strategy of "no screening" (Torrance et al., 1996, p. 65). This interpretation can be ambiguous as no screening generally does not necessarily represent no intervention at all as it usually represents no screening with therapy for symptomatically presenting disease. Irrespective of which interpretation is made, both conceptions of the ACER are not suitable for decision-making context (Siegel et al., 1996). It is important to estimate ICER correctly, especially as ICER and ACER estimates can be very different in the case of cancer screening (Neuhauser & Lewicki, 1975).

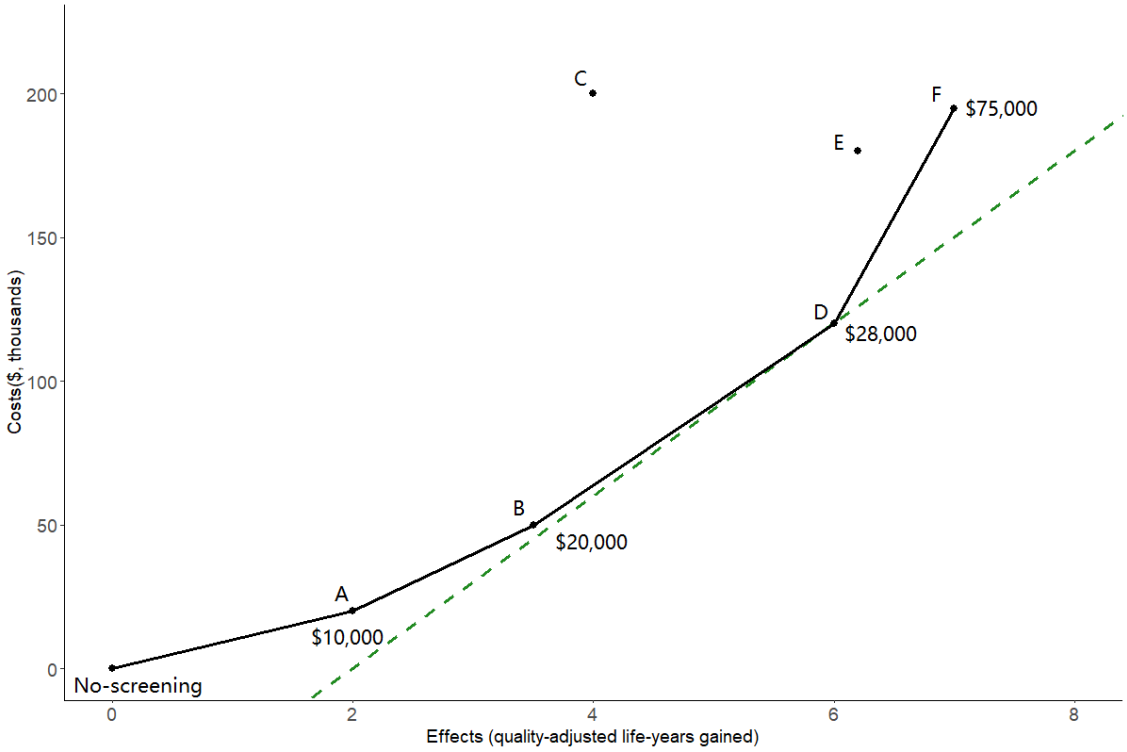
If a CEA included an inadequate range of strategies this can result in unreliable evidence for decision-making. Omitting relevant strategies as comparators can make any given strategy appear more cost-effective, which may lead policy makers to adopt a more than optimally intensive strategy. Any ICER that is estimated on the basis of a comparison set that omits relevant comparisons is not an adequate ICER. The ACER is a common special case where all strategies have been omitted and the strategy of interest is simply compared to no screening (or no screening and no treatment). It has long been recognised that appropriate ICER estimation should be based on the comparison of all relevant strategies (Torrance et al., 1996).

The cost-effectiveness plane assists in the visualisation of cost-effectiveness estimates (Black, 1990). The horizontal and vertical axis represent effectiveness and costs respectively. The origin of the cost-effectiveness plane is the net point for comparison. In the context of screening, the origin is often the no-screening scenario, i.e., treatment for symptomatic disease in the absence of screening. Linking non-dominated strategies on the cost-effectiveness plane visually presents the efficient frontier which is the combination of the strategies closest to the axis of effectiveness with the lowest costs. On the cost-effectiveness plane, the slope of the cord joining the estimates of cost and

health effects of two interventions on the efficient frontier is the ICER (Drummond et al., 2015, p. 56). The strategy on the efficient frontier with the highest ICER within the threshold identifies the most cost-effective screening strategy under a given decision threshold.

Figure 1-1 and Table 1-1 show a hypothetical example of decision-making with CEA evidence without considering uncertainty. In this example, Strategy C is subject to simple dominated by Strategy D and Strategy E is subject to extended dominated by a combination of Strategies D and F. After removing the strategies dominated, the efficient frontier has five screening strategies, namely no-screening, Strategy A, Strategy B, Strategy D, and Strategy F. Their corresponding ICERs are calculated based on pair-wise comparisons with this order. For example, the ICER of Strategy D is based on the comparison to Strategy B. With a decision threshold of $30,000 per QALY (green dashed line on Figure 1-1), Strategy D is the optimally cost-effective strategy among all the programmes.

Figure 1-1 A cost-effectiveness plane featuring ICERs, the cost-effectiveness threshold and the optimally cost-effective strategy (Strategy D)



The cost-effectiveness results listed in Table 1-1 are plotted on this cost-effectiveness plane. The solid black line represents the efficient frontier of this analysis, with corresponding incremental cost-effectiveness ratios labelled next to the strategies on the frontier. The green dashed line is a tangent line passing through the optimal strategy with a slope equal to the decision threshold of $30,000 per quality-adjusted life-year. Strategy C is strongly dominated and Strategy E is extendedly dominated by other strategies in this analysis.

Table 1-1 The calculation of ICERs

Strategy	Effects, (QALYs)	Costs ($, Thousands)	Incr. effects, (QALYs)	Incr. Costs, ($, thousands)	ICER ($/QALY)	ACER ($/QALY)
No-screening	0	0	-	-	-	-
A	2.0	20	2.0	20	10,000	10,000
B	3.5	50	1.5	30	20,000	14,286
C	4.0	200	-	-	SD	50,000
D	6.0	120	2.5	70	28,000	20,000
E	6.2	180	-	-	ED	29,032
F	7.0	195	1.0	75	75,000	27,857

ACER, average cost-effectiveness ratio; ICER, incremental cost-effectiveness ratio; Incr., incremental; QALYs, quality-adjusted life-years; SD, simple dominance or strong dominance; ED, extended dominance. † The cost-effectiveness results are plotted on this cost-effectiveness plane (Figure 1-1). After excluding the strategies that are SD or ED, the ICERs are calculated based on the comparison between the chosen strategy and the next more costly or effective strategy. The ACERs are estimated by comparing the strategy to the no-screening strategy. The optimal strategy D is chosen as it has the highest ICER below the decision threshold of $30,000 per QALY.

Guidelines for the choice of screening strategy in CEAs

Checklists for economic evaluation

In order to understand current recommendations regarding CEAs of screening this thesis reviewed existing checklists. The following material provides a specific focus on how the existing checklists suggest the choice of interventions, which is the main challenge in the CEAs of screening. This thesis is also interested in how they suggest the comparison among the interventions. Structured checklists for economic evaluation have been developed to ensure the quality of CEAs by examining the CEAs in a systematic and standard fashion. Commonly-used checklists include the Drummond Ten Point checklist (Drummond checklist), Drummond and Jefferson checklist, the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist, the Consensus on Health Economic Criteria (CHEC)-list, Philips checklist, and the Quality of Health Economic Studies (QHES) checklist (Frederix et al., 2015; Min et al., 2021; Watts & Li, 2019). Although these checklists can be applied to other forms of economic evaluation, such as cost-minimisation analysis (CMA), this thesis focuses on their application to CEA.

Drummond and Jefferson checklist (Drummond 35-point checklist)

This checklist was initially developed by Drummond and Jefferson for reviewing CEA submissions to the British Medical Journal (Drummond & Jefferson, 1996). It examines a total of 35 items from study design, data collection, to analysis and results interpretation. The assessment results for each item can be recorded as “yes”, “no”, or “not clear”. For some items, it has an additional option of “not appropriate”.

Importantly, this checklist is an assessment of reporting as opposed to the conduct of CEA. As long as the author of the CEA disclosed relevant information on the above items, CEAs met these criteria. Accordingly, an analyst might not apply methods appropriately but as long as the reporting of the methods applied is clear the study will score well against the checklist.

In this checklist, two of the items in the study design are relevant to the choice of comparators:

(4) The rationale for choosing the alternative programmes or interventions compared is stated.

(5) The alternatives being compared are clearly described.

Drummond and Jefferson (1996) suggested the considered comparator should be the existing most cost-effective alternative intervention or the one most widely used. They also mentioned that placebo is usually not the best comparator unless "do-nothing" is current practice.

Regarding the intervention comparison, incremental analysis is expected to be reported in the Drummond and Jefferson checklist:

(30) Relevant alternatives are compared.

(31) Incremental analysis is reported.

(32) Major outcomes are presented in a disaggregated as well as aggregated form.

This checklist suggests the incremental analysis should be reported based on the comparison to the relevant alternatives and the key components of cost and effectiveness should be reported in addition to the cost-effectiveness ratio (CER). For example, direct and indirect costs should be disclosed along with the overall cost. LYs gained with and without quality-of-life (QoL) adjustments should be also separately listed.

CHEERS checklist

The CHEERS checklist is published by a leading research society, ISPOR Task Force (Husereau et al., 2013). Like the Drummond and Jefferson checklist, the CHEERS checklist is explicitly a reporting checklist. The authors explicitly note that the checklist is not intended as a guideline for the conduct of CEAs. Again, this means that studies can meet the checklist items if they are clearly reported rather than necessarily being conducted appropriately.

The CHEERS checklist includes a total of 24 items and provides detailed explanations and examples of ideal disclosure for every assessment item. The items in the CHEERS checklist are in order of sections commonly used in a publication, from title, abstract, introduction, methods, results, discussion, and other issues. For some items, it has

additional recommendations for different study designs of CEAs. For example, recommendations on characterising uncertainty are separately described for single study-based and model-based economic evaluation.

Regarding the choice of comparators, the CHEERS checklist has one corresponding item:

> *(7) Describe the interventions or strategies being compared and state why they were chosen.*

In their explanation, it mentioned that CEAs should consider "do-nothing", "current practice", or "the most cost-effective alternative" as relevant comparators. The guidelines do not state that all potentially relevant comparators should be assessed but that authors should consider listing all potential alternatives and explain why comparators not assessed were excluded from the analysis.

The CHEERS checklist also has one item evaluating the incremental analysis:

> *(19) For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.*

The CHEERS checklist detailed how to calculate ICERs by emphasising pairwise comparisons and explained the circumstances that ICER is not applicable as intervention is dominant, dominated or not considered relevant for the context. This checklist did not explain how to choose the comparator applied to the formula for the ICER calculation. In general, this checklist focuses on if the CEA reports an ICER, rather than how ICER is calculated.

Drummond checklist

Drummond checklist is detailed in Chapter 3 of the textbook of Drummond et al., (2015). This textbook explained the elements of a sound economic evaluation with this checklist. It considers ten aspects, and each aspect has up to six supporting questions. The evaluation outcomes for each aspect are "yes", "no", or "can't tell". The textbook provides two assessment examples of Blomström et al. (2008) and McKenna et al. (2010). Unlike the Drummond and Jefferson and the CHEERS checklists, Drummond

checklist scrutinises the methods applied in the CEA, not just aspects of reporting. The guidelines in Drummond checklist are explicitly described as a guide for those wishing to appraise or conduct a CEA and as such different from the reporting nature of the previous guidelines. For example, the fourth checklist question addresses the identification of relevant costs and consequences (influencing the effectiveness), and the fifth and sixth questions relate to the valuation of these identified items.

The Drummond checklist examines the choice of comparators in the CEA:

> *(2) Was a comprehensive description of the competing alternatives given? (i.e. can you tell who did what to whom, where, and how often?)*
>
> > *(2.1) Were any relevant alternatives omitted?*
> >
> > *(2.2) Was (should) a 'do nothing' alternative (be) considered?*
> >
> > *(2.3) Were relevant alternatives identified for the patient subgroups?*

The Drummond textbook specified that the relevant comparators could be the intervention with different intensities and the various sequences of care pathways. The former can clearly be interpreted as requiring consideration of alternative screening frequencies and age ranges in the context of CEAs of screening. When considering the comparators, it is essential to consider the feasibility of "do-nothing" (Drummond et al., 2015). Also, the Drummond textbook reminds readers to check whether the decision problem is constrained by the omission of relevant comparators. Although the textbook noted the importance of minimising the risk of important comparator omission, it did not explain when and how the comparator omission happens.

The eighth item in the Drummond checklist and the first sub-questions to the tenth are relevant to ICER calculations:

> *(8) Was an incremental analysis of costs and consequences of alternatives performed?*
>
> > *(8.1) Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?*
>
> *(10) Did the presentation and discussion of study results include all issues of concern to users?*

(10.1) Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion?

Chapter 3 of Drummond textbook identifies the ICER between two interventions as the ratio to consider if the decision question is which intervention to be chosen rather than ratios based on comparisons to the do-nothing scenario. The textbook also suggests CEAs summarise their results with a CER. Although Drummond et al. (2015) is not definitive and explicit that the ICER must be used, it strongly implies this is what should be used. The ideal ICER calculation is detailed in other chapter of Drummond textbook, not in the introduction to the checklist (Drummond et al., 2015, pp. 98-102).

CHEC-list

The CHEC-list was proposed for methodological quality assessment in the systematic reviews of CEAs (Evers et al., 2005). There are 19 items included in the CHEC-list. Each item can be answered with "yes" or "no". This checklist does not provide details on the assessment items.

The following items in the CHEC-list are perhaps relevant to the comparator choice:

(2) Are competing alternatives clearly described?

(4) Is the economic study design appropriate to the stated objective?

The CHEC-list has one item relevant to the incremental analysis:

(13) Is an incremental analysis of costs and outcomes of alternatives performed?

It is difficult to interpret the aspects that these items intend to evaluate because the CHEC-list did not explain the criteria or provide any examples.

Philips checklist

Philips checklist focuses on the quality assessment of decision-analytic models in CEAs (Philips et al., 2006). This checklist evaluated the quality of models from three dimensions, namely *structure*, *data*, and *consistency*. These dimensions separately have nine, four, and two elements. A total of 15 items are assessed by this checklist. Only one

item is relevant to the comparator choice and none discussed the calculation of ICER. The answer can be “yes”, “no”, “not applicable”, or “unclear”.

The fifth item in the dimension of *structure* suggested the inclusion of comparators:

> *Is there a clear definition of the options under evaluation?*
> *Have all feasible and practical options been evaluated?*
> *Is there justification for the exclusion of feasible options?*

The Philips checklist suggests CEAs should consider all the possible strategies, including the comparator(s) representing the current practice. The screening strategies should not be restricted by data availability, constraints of current practice, or the decision-makers’ immediate concerns. Justification is required if any feasible strategy is excluded.

QHES checklist

The QHES checklist provides a quantitative measurement of the quality of CEAs (Chiou et al., 2003). This checklist evaluates 16 items with different weights, ranging from 1 to 9. This grading system has the sum of the weights to be 100. While Chiou et al. (2013) gives detailed explanation of the grading system it does not provide guidance regarding the evaluated items. The QHES checklist does not explicitly mention the issue of comparator choice. Only two items are probably relevant to the incremental analysis:

> *(6) Was incremental analysis performed between alternatives for resources and costs?*
> *(12) Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner?*

The QHES checklist listed criteria but did not explain their meaning or provide any examples. Without explanation, it is hard to judge what the "numerator" and "denominator" refer to and if the 12^{th} item is relevant to the ICER calculation.

None of the checklists is specifically designed for the CEAs of screening. The examples used these guidelines are CEAs of therapies. The issues of ambiguity and limited

suitability for screening applications of these guidelines might lead to difficulties when applying them in the context of screening. For example, both Drummond and Jefferson and the CHEERS checklists recommend that CEAs consider the existing most cost-effective alternative as the comparator. This is suitable in CEAs of therapies but not in screening as the optimally cost-effective screening strategy is typically unknown before the conduct of the simulation. Moreover, simply selecting the current screening strategy (if screening is provided) as the comparator does not ensure that the most relevant efficient comparator has been provided.

Textbooks and HTA guidelines

Some recommendations from other sources can also be used as a benchmark for quality assessment of CEA, such as textbooks of economic evaluation, and guidelines published by national health technology assessment (HTA) authorities and international research groups. In addition to the Drummond checklist which is detailed in the textbook of Drummond et al., (2015), other textbooks also play an important role in the guidance of conducting CEAs. For example, Torrance et al. (1996) in the textbooks of Gold et al. (1996) explained that CEAs should consider a range of possible screening intensities as comparators. While Torrance et al. (1996) gave clear screening-specific guidance, the checklist of Siegel et al. (1996) given in Gold et al. (1996) is not very informative. Siegel et al. (1996) did not detail comparator inclusion in their proposed checklist and this checklist focuses on reporting only. Similarly, as a reporting checklist, the updated checklist of Gold et al. (1996) does not examine the comparator choice, but notes the reporting of other intervention descriptors including the intensity and timing of intervention (Neumann et al., 2016).

The guidelines issued by HTA authorities are to ensure the methodological quality of CEAs and ensure the comparability of CEAs estimates between analyses for their jurisdictions. This thesis reviewed some well-known examples of such guidelines, including guidelines of the Health Information and Quality Authority (HIQA) in Ireland (HIQA, 2020), the National Institute for Health and Care Excellence (NICE) in England and Wales (NICE, 2022), the National Health Care Institute (Zorginstituut Nederland, ZIN) in Netherlands (ZIN, 2016), Canadian Agency for Drugs and Technologies in Health (CADTH) (CADTH, 2017), and the Pharmaceutical Benefits Advisory Committee (PBAC) for Australian Government Department of Health (Australian

Government Department of Health, 2016). Notably, this thesis did not find any CEA guidance published by the UK National Screening Committee.

All the HTA authorities-published guidelines describe the consideration of current care as comparators in CEAs (Australian Government Department of Health, 2016; CADTH, 2017; HIQA, 2020; NICE, 2022; ZIN, 2016). ZIN mentions standard of care and/ or usual care should be included (ZIN, 2016). Standard of care is the intervention recommended as the first choice in the clinical guidelines and usual care is the care procedure routinely implemented in clinical practice (ZIN, 2016). NICE specifies the relevant comparators are the established practice in the National Health Service (NHS). HIQA recommends using routine care, which is widely used in clinical practice, as the preferred comparator (HIQA, 2020). CADTH suggests considering the interventions that are currently used, potentially displaced, and nearly available (CADTH, 2017). Similarly, the PBAC guidelines suggest the current alternatives and the therapies most likely to be replaced, including a medicine and care listed in Pharmaceutical Benefits Scheme or not (Australian Government Department of Health, 2016).

Among all the guidelines, only HIQA's gives suggestions on the choice of comparators in the context of screening programmes:

> *In the evaluation of public health interventions there may be scope to define a wide range of comparators that are different configurations of the same basic intervention. For example, a screening programme based on a particular diagnostic test may be specified for different age ranges and screening frequencies, potentially generating a very large number of comparators. Omission of potential comparators can impact on the estimated cost-effectiveness of included interventions. In these cases, justification should be given for the included and excluded comparators, preferably with reference to their clinical plausibility and organisational feasibility. (HIQA, 2020, p. 26)*

Although HIQA recognises the technical difficulty of simulating such a huge number of comparators and allowed CEAs to limit the number of comparators in the context of therapeutic products, HIQA does not suggest how to deal with a huge amount of

screening comparators (HIQA, 2020). Other HTA guidance does not explicitly mention screening, but that might be expected as screening does not fall within their remit.

NICE guidance does not explicitly mention screening, but mentions the "diagnostic technology" in the HTA document which can refer to the screening modality (NICE, 2022). NICE recommends that CEAs identify all relevant comparators in practice, and that when considering a diagnostic technology's care pathway, CEAs include any variations based on test results or technologies used (NICE, 2022). Diagnostic sequences are also mentioned in the NICE guideline for describing care pathways (NICE, 2022).

CADTH (2017) clearly documented, "all comparators should be considered and the choice to remove any comparator(s) from the decision problem clearly justified." CADTH recommends the choice of comparators should be conceptual-driven, not based on the available data (CADTH, 2017). Although CADTH (2017) does not establish guidelines for screening programmes directly, management strategies are mentioned in existing CADTH guidelines. The combination of interventions, such as the use of a screening test and a treatment, is one of the examples of management strategies in CADTH guidelines (CADTH, 2017). CADTH (2017) suggests CEAs should consider possible combinations of interventions and the sequence of implementation of these interventions if the interventions are applied sequentially.

ZIN (2016) does not offer guidelines for the choice of comparators in the context of screening programmes. Compared to other HTA documents, ZIN (2016) emphasizes the use of recent national and international guidelines for underlying the comparators in the CEAs. PBAC guidelines did not provide additional information to the CEAs of screening, although CEAs of fixed-dose combination products, nutritional products, vaccine products, and codependent technologies have additional guides for submission to PBAC (Australian Government Department of Health, 2016).

International research groups and agencies also publish recommendations on how to conduct a CEA. For instance, ISPOR publishes a series of guidelines for methodological suggestions in economic evaluation and WHO produces its guidelines which often apply to low and middle income countries (LMICs) (WHO et al., 2003).

Possible variations in screening strategies

Depending on the availability of good data, CEAs can be grouped into “what-is” or what-if” studies (Owens et al., 2016). Unlike the “what-is” studies which have reasonably good data available to support the estimates of cost and outcome, "what-if” studies usually do not have good quality data available but are required as they address relevant policy questions before data becomes available (Owens et al., 2016). A typical example of "what-if” studies is the CEA of screening, in which analysts can evaluate interventions with different starting age and stop age or different target groups without direct evidence (de Koning et al., 2014; Owens et al., 2016).

Many checklists mention the consideration of “do-nothing”, “current practice”, and “the most cost-effective alternative” for the choice of comparators. In theory, CEAs should identify and include all the possible comparators (Drummond et al., 2015, p. 100; Torrance et al., 1996). It is important to include strategies with different intensities as comparators in the screening CEAs (Torrance et al., 1996, p. 65). Simple forms of variation in screening intensity include variation in either or both the screening frequency and age ranges (Neumann et al., 2016, p. 87). For example, the variation of disease risk with age not only provides the rationale for when to start and stop screening but it may also provide a rationale for more frequent screening within certain ages, such as the shorter screening interval often offered to women under age 45 relative to older women within cervical screening programmes. Another possible variation to consider is risk factors within age groups (Neumann et al., 2016, p. 87). The heterogeneity of screening consequences may lead to different optimal strategies for different subgroups (Neumann et al., 2016, p. 87).

The strategies can be any variety in terms of screening primary tests and conditional sequential tests, in addition to the possible combinations of therapeutic options conditional on screening results (Drummond et al., 2015; Neumann et al., 2016, p. 153). Screening programme typically comprises a primary screening test and triage test(s). A common triage example in cervical cancer screening is applying pap cytology to individuals with human papillomavirus (HPV)-positive and individuals with abnormal cytology results then have colposcopy (Stoler et al., 2020). The triage algorithms can be long and detailed (Stoler et al., 2020).

Screening CEAs face a challenge in considering all the possible strategies (Drummond et al., 2015, pp. 23-24). There may be important practical constraints on simulating many strategies. Some models might require specific code to simulate each strategy and thus the construction of a model capable of simulating multiple strategies might be time-consuming. Even when a model has a sufficiently flexible structure that does not require extensive code to add additional strategies, the fact of simulating multiple strategies can require much longer run times that might be required by a more conservative number of strategies.

Economic evaluation helps minimise the chances of failure to include an important strategy for consideration and the research question should not be restricted by the existing effectiveness evidence or by the person commissioning the CEA (Drummond et al., 2015, p. 45). Though clearly there could be unintended consequences if such a narrowing of the research question led to the omission of relevant comparators. Comparison among a limited number of strategies leads to a biased estimation of ICER, which influences policy choice in decision-making. For example, some CEAs of HPV-based screening does not consider cytology which is also available as the primary screening modality for cervical cancer (Campos et al., 2015; Campos et al., 2012; Shi et al., 2011).

The existing literature does not provide much guidance on the challenges related to a large number of potential strategies. Drummond et al. (2015) suggest limiting the number of strategies considered by removing those that have a small chance of being cost-effective. Torrance et al. (1996) recommend only including the screening strategies that intensities are really feasible in reality, stating “The analyst will have to decide at what point comparisons are no longer realistic and when programs differ little enough in their effect that finer distinctions will not offer much additional insight” (p. 66). While both suggestions are sensible neither offer explicit guidance on exactly how to select strategies and rely on a degree of subjective judgement on the part of the modeller. We require further research to provide detailed instructions on how to choose the strategies in the CEAs of screening.

Recent systematic reviews of cancer screening

This section reviews the choice of screening strategies in the CEAs of cancer screening with existing systematic reviews. Some reviews have explicitly evaluated the range of strategies included in the CEAs of screening. Although other reviews did not note this issue explicitly, they summarised the inconsistent inclusion of screening strategies by modalities, age ranges, and intervals. From their observations, CEAs of cancer screening generally do not consider enough range of screening strategies.

Existing evidence has shown many CEAs of cancer screening have biased ICER estimates due to inadequate comparator inclusion and inaccurate comparison. Fabbro et al. (2022) conducted a systematic review of lung cancer screening and found over half of CEAs were published since 2018. Of all the CEAs included, 70% (n = 23) only considered a single screening interval and almost half (48%, n = 16) actually presented ACERs instead of ICERs. A similar observation on comparator omission is found in the CEAs of cervical cancer screening. O'Mahony et al. (2015) estimated 70% (n = 21) of CEAs omitting relevant comparators by observing the ICERs of the identified optimal strategy and the next efficient strategy. A similar observation is also found in the literature of breast cancer screening (O'Mahony, 2014, 2019). These reviews pointed out the issues existing in the credibility of ICER estimates in the CEAs of cancer screening.

Other systematic reviews indirectly demonstrated possible comparator omission in the CEAs of cancer screening. A systematic review of CEA of oral cancer by Thankappan et al. (2021) found the reviewed CEAs typically did not consider a large array of strategies with eight being the most assessed within the sample. This leaves it highly likely that the studies did not provide a sufficient investigation of possible screening strategies. Regarding the screening age, Rashidian et al. (2013) found less than half of CEAs of breast cancer screening (n = 11, 39%) vary the screening age range. Most (n = 21, 75%) considered screening strategies aged either below 50 (n = 2, 7%), 50-70 (n = 18, 64%), or over 70 (n = 1, 4%). From their summary of screening strategies, it is observed that more than half of CEAs only studied a single screening interval. A review of economics analyses of prostate cancer screening by Sanghera et al. (2018) shows CEAs tend to include a limited range of screening intervals (Sanghera et al., 2018). This review found 90% (n = 9) CEAs ignored the availability of five-yearly screening, while they considered

annual (n = 5), biennial (n = 6), and quadrennial (n = 7) screening. These examples show that existing CEAs of screening demonstrate inadequate inclusion of strategies across a range of different cancers.

Many systematic reviews of CRC screening have been conducted but only one has considered the issue of comparator omission in the reviewed CEAs. Systematic reviews evaluated the CEAs of CRC screening with the CHEERS checklist, only focus on disclosure of methods applied (Gheysariyeha et al., 2022; Mendivil et al., 2019; Zhong et al., 2020). Hanly et al. (2012) only found one CEA that failed to state the rationale for the choice of strategies simulated based on the Drummond and Jefferson checklist, though this is simply a critique of the reporting and not the underlying choice of strategies compared. The majority of systematic reviews of CEAs of CRC screening use existing checklists which are not suitable for examining the comparators.

The question of the inclusion of sufficient strategies has only been raised in one of the reviews of CEAs of CRC screening surveyed. Using the Philips checklist, Jeong and Cairns noted that the intervals considered were not sufficient, though they did not give any detail on why they considered the strategies insufficient and the issue only arose within the specific context of computed tomography colonography (CTC) (Jeong & Cairns, 2013). A more general review of the choice of strategies considering the more commonly used modalities of stool and conventional colonoscopy is therefore needed.

The correct calculation of ICERs is another methodological issue in CEAs of CRC screening that merits consideration, especially the confusion of ICERs and ACERs. Existing reviews with quality checklists cannot properly evaluate the accuracy of ICERs interpretation because the checklists only request incremental analysis and do not clearly address this issue of the correct basis of comparison between strategies. The systematic reviews of Pignone et al. (2002), Lansdorp-Vogelaar et al. (2011), Patel and Kilgore (2015), and Khalili et al. (2020) separately summarised the cost-effectiveness results based on the comparison to the no-screening and the ICERs based on the comparisons among all included modalities. While both results are presented with CERs, they have different meanings and need to be interpreted carefully.

The use of terminology should be precise when presenting ACERs and ICERs to ensure their correct interpretation. Pignone et al. (2002) and Lansdorp-Vogelaar et al. (2011) explained how the correct ICER was calculated in their methods sections, including definitions of weak and strong dominance. Notably, both reviews did not use the term "ICER" when presenting the results compared to the no-screening. Instead of "ICER", Pignone et al. (2002) mainly used ACER and CER, and Lansdorp-Vogelaar et al. (2011) only used CER. Lansdorp-Vogelaar et al. (2011) noted some CEAs reported ICERs based on the comparison to the no-screening, causing the difference between calculated ICERs and reported ICERs. The systematic review of Patel and Kilgore (2015) and Khalili et al. (2020), however, both inappropriately used the term "ICER", which ICERs can be based on the comparison to the no-screening or the intervention representing the current standard of care. This use of ICER for non-ICER comparisons creates scope for confusion and sub-optimal resource allocation.

In summary, existing reviews of screening for cancers other than CRC indicate strategy selection is an area of weakness in many applied analyses. The issue has received much less attention in the CRC context. The failure of existing reviews to identify the issue clearly is likely a consequence of shortcomings in checklists regarding strategy selection and ICER comparisons. There is a gap in the literature to determine the extent of the problem of inappropriate strategy selection and comparison in CEAs of CRC screening.

Part 2 Cost-effectiveness models applied in cancer screening

CEA modelling is a tool to simulate the anticipated costs and effects of screening programmes (Mandelblatt et al., 1996, p. 160). Some models have computational constraints regarding the number of strategies they can simulate which is a possible reason for CEAs considering a limited range of screening strategies. These constraints may be due to either or both structural constraints within models that make simulating many strategies difficult and the simple issue that increasing the number of simulated strategies will increase simulation runtimes, which can present a meaningful problem for slower models. In order to provide context to a possible solution to the omission of relevant strategies, this section is going to review the current simulation methods applied in the CEAs of screening. This section will introduce common model types applied in

cancer screening CEAs, open-source models and tutorials, and modelling software and tools.

Types of decision-analytic models

Several types of decision models are applied in CEA, including decision trees, cohort state-transition models, microsimulation models, dynamic transmission models, and dynamic simulation models (Neumann et al., 2016). A modelling framework should be chosen based on the attributes of the decision problem. The considerations include how to manage time and the unit of analysis (individual or cohort) in the model and whether individuals interact with each other or the model (Neumann et al., 2016, p. 117). Notably, the model types are not always mutually exclusive. For example, microsimulation can refer to an individual-level discrete time state-transition model or a discrete event simulation (DES) which is one form of dynamic simulation models (Standfield et al., 2014).

Several types of simulation models are commonly applied in CEAs of cancer screening. A recent systematic review of cervical cancer screening CEAs found half (n = 19) used Markov models (Viscondi et al., 2018). Another review of cervical cancer screening CEAs found of the 19 CEAs, 11 (58%) used individual microsimulation models, 7 (37%) used Markov models, and only a single study (5%) applied a decision tree model (Mezei et al., 2017). A systematic review of breast cancer screening summarised the most common model type is cohort state-transition model (n = 27, 77%), followed by decision tree (n = 2, 6%) and DES model (n = 2, 6%) (Schiller-Frühwirth et al., 2017). CEAs of CRC screening tend to use Markov models (n = 14, 88%) and microsimulation models (n = 2, 13%) (Hanly et al., 2012). Among 33 CEAs of CRC screening, Markov models are the most frequently used (n = 23, 67%) and microsimulation (n = 10, 30%) is found in the remaining CEAs (Ran et al., 2019). The following introduction are the model types commonly applied to screening CEAs, including decision tree, state-transition model, and DES:

Decision trees

Decision trees are a method to structure the problem and calculate the expected cost and effectiveness of a specific intervention by considering the probability of each event in

the model (Karnon & Brown, 1998). This type of model incorporates the likelihood of a series of events happening.

The primary advantage of decision trees is the accessible visual representation of a decision problem. These models demonstrate the comparators and their courses of consequences. Decision trees are suitable when the decision process can be easily identified in a conditional control flowchart. This type of model, however, is not an appropriate choice when the decision process is complex, due to a large number of comparators, events, and outcomes (Drummond et al., 2015, p. 331). Also, decision trees do not suit a long-time frame or a recurrence of important events (Kuntz et al., 2016, p. 118). As such, decision trees are not a natural model choice for cancer screening.

Discrete-time state-transition models

State-transition models are suitable for decision problems that involve changing clinical health states over time (Siebert et al., 2012). State-transition models simulate membership of health states and the transitions between them. A common type of such models is the discrete-time state-transition model which divides the time horizon (model duration) into equal increments of time which is the cycle length of the model. This cycle length is typically fixed throughout the simulation. A state-transition model is similar to a recursive decision tree (Jordan et al., 1996). One important assumption in state-transition models is that all individuals must exist in one of the discrete health states. Absorbing states are the health states that individuals cannot move to another state once they enter this health state. Death is a typical example of an absorbing state (Drummond et al., 2015).

State-transition models can be applied to single simulated individuals or, commonly, to groups or cohorts of patients. One commonly applied form of cohort-based state-transition model is known as a "Markov" model (Kuntz et al., 2016, p. 119). Markov models are appropriate to solve a decision problem that has a continuous risk over time, where events happen more than once, and the timing of events of interest has an important impact (Sonnenberg & Beck, 1993). A well-known weakness of Markov models is their "memorylessness". All entities are assumed to have the same transition probability in a given health state, regardless of their previous histories (the "Markovian assumption")

(Briggs & Sculpher, 1998). This aspect can be worked around by adapting the model structure to reflect prior events, but this comes at the cost of additional model complexity.

Markov models can be divided into two types: Markov chains and Markov processes (Drummond et al., 2015, p. 333; Wan et al., 2016). The main difference between these two types is the transition probability from one health state to another is assumed to be independent of time in Markov chains and can vary with time in Markov process (Alagoz et al., 2010; Karnon, 2003). It might be unrealistic to implement Markov chains when the risk of death or other transitions is closely associated with age. Conversely, although Markov processes may involve more plausible assumptions, they can be more difficult to implement.

Some challenges can occur when building a Markov model. As mentioned above, if the transition probabilities are time-variant, the model structure becomes more complex. In addition, simulation of a large number of health states and interventions is quite demanding. Although the restriction of Markovian assumption can be overcome by adding more health states, it results in a very large model or "state explosion" that is hard to manage (Siebert et al., 2012).

As noted above, state-transition models can be implemented at an individual level. This type of simulation is also called "first-order Monte Carlo" which draws the probability distribution of events for each individual (Weinstein, 2006). The mean outcomes from a large number of simulated subjects can represent the cohort outcomes of interest. An advantage of such a microsimulation approach is the ability to capture the patient-level variance, in contrast to the second-order uncertainty, parameter uncertainty (Groot Koerkamp et al., 2010).

DES

DES is one of the common individual-based simulation models (Marshall et al., 2015; Neumann et al., 2016, p. 120). Dynamic simulation model is suitable for the decision question that involves the consideration of interactions between individuals or resources in the healthcare system (Neumann et al., 2016, p. 118). DES has a feature of discrete events in which health states or events of interest are mutually exclusive (Karnon et al., 2012). DES simulates the sojourn time until the next event happens under a continuous

simulation clock time (Drummond et al., 2015, p. 337; Karnon et al., 2012). This means that individuals spend different lengths of time that can vary continuously between moving from one health state to another state. This differs from the discrete time approach applied in models such as Markov models in which the change in state membership is considered over discrete time intervals. The target simulated individuals, "entities", can be any interested people or objects, typically patients (Karnon et al., 2012).

DES has two main classifications, non- resource-constrained- model and resource-constrained- model (Karnon et al., 2012). Non-constrained-resource models are easier to implement but cannot reflect the capacity limitation, e.g., a limited amount of colonoscopies available for CRC screening. This capacity issue, however, can be eased by expanding the service or resources. In contrast, constrained-resource modelling is more complex but incorporates capacity limitation. This is well-suited for designing the procedure of a specific service that has limited resources because it considers the number of entities waiting in queues and their waiting time.

DES has the advantage of flexible modelling which can implement complex models and may produce more valid cost-effectiveness results (Karnon & Haji Ali Afzali, 2014; Simpson et al., 2009; Standfield et al., 2014). DES handles time continuously, permitting a simulation more representative of disease history (Karnon & Haji Ali Afzali, 2014). The flexibility of modelling makes DES an intuitive modelling approach and can be more computational efficient because DES only involves operations as they occur (Karnon et al., 2012). DES has a particular advantage in a screening context in which disease processes typically apply to only a minority within the simulation. As a microsimulation, DES permits transitions based on the attributes of entities, which overcomes the feature of memorylessness of Markov models (Simpson et al., 2009). Furthermore, microsimulation is easier to calibrate the model with data of heterogeneous subgroups (Karnon & Haji Ali Afzali, 2014).

Generally, the main difference among a decision tree, a state-transition model, and a DES is the way of representing the state transitions. A decision tree simulates the probability of the occurrence of an event without the explicit consideration of time or event recurrence for a cohort. A state-transition model simulates the health states and the probability of moving to another health state either at the cohort or individual level, often

simulated in discrete time. A DES describes the health states and simulates the individual's sojourn time from one state to another in continuous time. These model types have their own advantages and attributes making them naturally more presentable and suitable for some decision problems than other model types (Caro et al., 2012).

Open-source models

Open-source CEA models are desirable for enhancing transparency (Cohen et al., 2017; Eddy et al., 2012). Open-source is a description of software in which the original source code is openly available and modifiable. The concept can be applied to simulation models in healthcare in which the model code is made available to all wishing to use it without proprietary restrictions.

In the recent decade, journals or research communities started advocating for researchers to disclose the details of models to scrutiny (Eddy et al., 2012; Poole et al., 2007; Schramm et al., 2016). Open-source models have several advantages. A survey by Dunlop et al. (2017) investigated health economic stakeholders for their opinions on open-source health economic models. This survey found the most frequently mentioned benefit of open-source models is to encourage "trustworthy, reproducible, validated, comparable, and flexible health economic models". Also, this survey found the main reason for accessing models is "to learn technical aspects of the model for use in a different disease area or decision problem". Open-source models can improve the research quality and teach the application of modelling techniques.

Open-source models designed for treatments

There are several open-source models available in CEA, most of which are not designed for screening applications. An example is an open-source model built for pain therapy from the results of a systematic review (Sullivan et al., 2016). This model is a Markov model initially established in Microsoft Excel and replicated into the statistical program R. This model uses seven states to describe treatment for up to the third-line treatment. Another open-source model example is a model of Alzheimer's disease developed by the International Pharmaco-Economic Collaboration on Alzheimer's Disease (IPECAD) to evaluate the cost-effectiveness of a hypothetical disease-modifying therapy (Green et al., 2019). This is also a Markov model programmed in Excel and R. It has a total of 56 states

to depict the disease progression among predementia, Alzheimer's disease, and death. The PROSIT Disease Modelling Community, an international research project, also proposed an open-source Markov model for diabetes mellitus with six health states (Schramm et al., 2016). This model is built in spreadsheet software OpenOffice.org, which is free of charge online.

In 2017, the Innovation and Value Initiative (IVI), a non-profit research organisation, started the Open-Source Value Project (OSVP) which aims to improve credibility of decision-making in the US by providing flexible open-source models (Jansen et al., 2019). The IVI builds their models based on public comments, iteratively collecting the opinions from diverse stakeholders in their model development (Incerti, Curtis, et al., 2019). So far, the IVI has released two models, separately for rheumatoid arthritis (IVI-RA) and epidermal growth factor receptor-positive non-small cell lung cancer (IVI-NSCLC), and is currently preparing the model for major depressive disorder.

Both IVI-RA and IVI-NSCLC models are published in the format of R packages, the complete code for which can be found on GitHub. It has two versions of web application tools for interactive models. First, the Basic Value Tool is suitable for those not expert in the methodology of health economics. Second, the Advanced Value Tool is designed for those familiar with the methodology but do not want to inspect the R code. The Advanced Value Tool has higher flexibility in adjusting the model parameters and treatments to perform specialised CEA and multi-criteria decision analyses (MCDA).

The IVI-RA model, programmed in R and C++, was initially released in 2017 (version 1) and updated in 2019 (version 2) to incorporate the opinions collected from the public. The IVI-RA model is a discrete-time individual patient simulation designed for comparing the treatment strategies for patients with RA, in which different treatment sequences are allowed as comparators (Incerti, Curtis, et al., 2019). Conversely, the IVI-NSCLC model is an individual-level continuous-time state transition model published in 2019. The IVI-NSCLC model is designed for simulating alternative sequential treatment strategies for lung cancer. The IVI-NSCLC model has n + 2 health states for n treatments, to describe health states for each treatment line, a health state after all the treatments, and death.

Open-source models designed for cancer screening

This thesis only found two open-source models applied in cancer screening, separately for colon and prostate cancer. Both open-source models are detailed but do not represent an easily accessible example for junior analysts to use as training resources.

The open-source Prostata model is a microsimulation model for prostate cancer screening (Karlsson et al., 2019). This model was adapted from an existing US model (Etzioni et al., 2008; Gulati et al., 2013; Gulati et al., 2010) developed by the Fred Hutchinson Cancer Research Centre (FHCRC) to the setting of Sweden in 2019 (Karlsson et al., 2019). While the Swedish Prostata model was extended and adapted again to the UK context (Keeney et al., 2022), the Swedish version of Prostata model is the only model fully accessible on GitHub, coded in R and C++. This Swedish Prostata model is provided in the format of R package. Despite the FHCRC model built in C under an open-source licence, this code does not appear to be available online.

The Colon Modelling Open Simulation Tool (CMOST) is a framework used to estimate the efficiency and costs of CRC strategies (Prakash et al., 2017). This framework is designed to simulate the natural history of CRC. The CMOST is validated by comparing results to the randomised CRC screening trials and other existing CRC simulation models, including Microsimulation Screening Analysis (MISCAN) (van Hees et al., 2014), Colorectal Cancer Simulated Population Model for Incidence and Natural History (CRC-SPIN) (Rutter & Savarino, 2010), and Simulation Model of Colorectal Cancer (SimCRC) (Knudsen et al., 2010). The full source code is programmed in Matlab and available on GitLab. This framework has a graphical user interface (GUI) to define model inputs and run the simulation without inspecting the detailed code. The results can be saved in the format of PDF, Excel, and Matlab files.

The CMOST framework can apply seven screening interventions at one time. A range of screening modalities can be chosen, including colonoscopy, rectosigmoidoscopy, faecal occult blood test (FOBT), faecal immunochemical test (FIT), and DNA test (Septin-9). It allows adjustment of the test sensitivity, the test specificity, the fraction of the population choosing a given screening test, and the adherence to main screening tests and follow-up investigations. Users can define the screening programmes by choosing the starting year, the stop year, and the screening interval, e.g., every one, two, or five years.

Teaching models and tutorials

The models described above are applied models in that they are used to answer actual policy questions using data that is intended to be representative of reality. Despite an increasing number of applied open-source models being published, few models intended primarily as teaching tools are available. Teaching models can be used as a tool for new scholars or students to learn the methodology of economic evaluation. Most of the teaching models, however, are designed for discrete time state-transition modelling. This thesis only found one tutorial designed for DES (Davis et al., 2014).

Green et al. (2022) published a tutorial which is designed for Excel users to adapt to R. This tutorial explained the R code step-by-step by translating a three-state time-dependent Markov model example in Excel to R (Green et al., 2022). In addition to the basic Markov simulation, this tutorial also includes the code for ICER calculation, plotting a cost-effectiveness plane, and conducting a probabilistic sensitivity analysis (PSA). Their R code can be found on GitHub.

Unlike Green et al. (2022) having the transition probabilities dependent on the age of the study cohort, another tutorial published by Williams et al., (2017) provided a detailed parametric semi-Markov model which measuring the time since entering the current health state, instead of the initial state (Williams et al., 2017). This tutorial separates the steps of modelling into a series of functions which are available online. This tutorial also provides material for PSA and visualising the cost-effectiveness results with cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC).

The Decision Analysis in R for Technologies in Health (DARTH) international research group published a range of open-source tutorials illustrating how to build a model for CEA in R and apply them to demonstrate the methodology in economic evaluation. The model tutorials published by DARTH focus on discrete-time state-transition models. The R code can be downloaded in their appendices of publications and on GitHub.

In DARTH's tutorial on the individual state-transition model, a step-by-step guide was provided with an application of a hypothetical disease (the Sick-Sicker model) as an example (Krijkamp et al., 2018). The Sick-Sicker model, a simple Markov model, was first described to demonstrate the approach of best-fitting model inputs, which has only

four health states, namely healthy, sick, sicker, and dead (Enns et al., 2015). In this tutorial, the DARTH group also provided an extended Sick-Sicker model to incorporate "memory", allowing the mortality rate and the treatment effectiveness to be dependent on an individual's sojourn time of the sick states and its baseline characteristics (Krijkamp et al., 2018).

DARTH also published tutorials to guide to building cohort state-transition models and coping with some limitations of traditional Markov models. They proposed a multidimensional dynamics-array approach to save the details of transitions among health states (Krijkamp et al., 2020). This modelling approach is faster and requires less memory than the cohort trace approach if a large number of health states are modelled. In addition, DARTH published a tutorial on time-dependent cohort state-transition models (Alarid-Escudero et al., 2022). This tutorial shows Markov processes which incorporate the transition probabilities dependent on the simulation time or state-residence time. DARTH's tutorials attempt to overcome the challenges, the "Markovian assumption", as implementing state-transition models.

We are aware of only one open-source DES tutorial model. The NICE Decision Support Unit produced a technical support document for illustrating how to build a DES in Visual Basic for Applications (VBA) in Excel, R, TreeAge Pro, and SIMUL8 (Davis et al., 2014). This tutorial applied an illustrative example of osteoporosis treatments and the code can be downloaded online.

To the best of our knowledge there is no open-source teaching model designed for screening interventions, irrespective of model type. While there are open-source screening models applied to screening, these applied models are complex and are thus not suitable as teaching examples. More generally, there remain very few tutorials designed for DES.

Modelling software programmes

Many software or computer programming languages can implement CEA models. Common simulation software often uses a "point-and-click" user interface and has some support programming languages for extension, such as TreeAge, Excel, Arena and SIMUL8. Conversely, programming languages only rely on the user-typed commands,

including R, Python, Matlab, Julia, Fortran, and C++. There is no consensus on which software or programming languages should be formally accepted or refused for governmental technology appraisals (Tosh & Wailoo, 2008).

These tools have their own advantages and disadvantages, making some more frequently used in the CEAs. For example, general purpose programming languages, such as R and Matlab, have flexibility in modelling but require coding skills and experience (Hollman et al., 2017). This takes time for an individual researcher to acquire and usually only large research groups have the supports to build a model with programming languages. Because rapid operation of a model may require more in-depth coding skills to operate models such as parallelizing a model and using C++ (Hollman et al., 2017; Incerti & Jansen, 2021), such complex tasks are typically not feasible for many research units or junior researchers starting off.

Excel, TreeAge, and R are commonly used for NICE technology appraisals (Tosh & Wailoo, 2008). In the survey of Tosh and Wailoo (2008), Excel is the most common software and every respondent uses Excel to build CEA models. While TreeAge is the easiest to be implemented and produces results diagrams efficiently, the acquisition cost of software hinders access (Menn & Holle, 2009). Although Excel is proprietary, it is so ubiquitous that there is little meaningful software acquisition cost barrier (Menn & Holle, 2009). Its ubiquity means that many researchers will have at least the basic skills of operation and are familiar with the spreadsheet environment. In addition, the widespread use of Excel means there is an abundance of teaching materials to support its use (Briggs et al., 2006; Davis et al., 2014; Gray et al., 2010).

Despite their popularity, Tosh and Wailoo (2008) noted Excel and TreeAge may not be capable of complex modelling. Although it is simple to learn Excel, even modestly complex models typically require the VBA extension and attendant coding capabilities, which largely increases its difficulty (Hollman et al., 2017; Menn & Holle, 2009). Regarding transparency and validation, “cell chasing” can be time-consuming and challenging in Excel (Tosh & Wailoo, 2008). It has been suggested that it may be more worthwhile spending time learning R, rather than VBA in Excel, considering R has more advantages than Excel (Hollman et al., 2017). R has another advantage of productivity

that can perform statistical analysis while TreeAge cannot and Excel is not sophisticated in that respect (Hollman et al., 2017).

R is a programming language and environment for statistical analyses and graphics. R has statistical techniques implemented and can be extended easily by creating "packages". Over 8,000 community-developed open-source packages are available which can be downloaded via the Comprehensive R Archive Network (CRAN) (Hollman et al., 2017). CRAN is a collection of sites around the world that carry identical R distributions, extensions, documentation, and binaries (Hornik & R Core Team, 2022). Users can access the latest official release of R preferably via the closest CRAN site to reduce network load (Hornik & R Core Team, 2022). In addition to CRAN, other repositories for storing packages include GitHub, R-Forge, and Bioconductor (Jalal et al., 2017). Furthermore, for computationally-intensive tasks, advanced users can incorporate other programming languages, including C, C++, and Fortran, to R projects.

Using R for CEA models has many advantages. First, it is free of charge. Second, R is flexible and multifunction that can simulate diverse model structures and perform statistical analysis (Hollman et al., 2017; Incerti, Thom, et al., 2019). Third, R is script-based which is easier to validate and has a high degree of transparency (Hollman et al., 2017). Fourth, R has a comparatively low processing time, allowing R to simulate a large number of simulations, such as the expected value of partially perfect information (EVPPI) analysis (Hollman et al., 2017). This, however, can face constraints eventually if the modeller is not proficient at C++. Fifth, R appears to be more frequently used in the recent health decision literature (Jalal et al., 2017) and R models have been accepted by NICE (NICE, 2022). Last but most importantly, R is supported by its large network of users and their open-source codes (Jalal et al., 2017). Incerti et al., (2019) summarised extensive R packages that can support economic evaluation.

Despite the advantages listed above, there are some disadvantages for R applied to CEA models. First, R requires more skills to implement and is more difficult to learn compared to other commercial software (Hollman et al., 2017). As a programming language, learning programming is required. Typed commands can be intimidating to many. Second, package dependencies and version control can introduce compatibility issues (Jalal et al., 2017). Version conflict or package broken might inhibit transparency and

have reliability concerns. Third, R is also not yet widely accepted by regulatory bodies partly as the HTA community still lacks experience in R (Incerti, Thom, et al., 2019). Another potential reason is that there is no commercial indemnification for the software and its packages. A package's programming error can be propagated to all the analyses employing this package (Incerti, Thom, et al., 2019).

In summary, R has an important role in decision modelling. Of its flexibility, R can overcome some of the modelling constraints in Excel. Research groups advocate using R for decision models and we can see a trend toward R modelling (Incerti, Thom, et al., 2019; Jalal et al., 2017). More publications and tutorials on R are desired.

Part 3 CEAs considering thousands of screening strategies

Currently, there is no clear checklist or any structured criteria regarding how to judge if a CEA has considered enough relevant strategies as comparators for the resulting cost-effectiveness estimates to be considered a reliable guide to policy. Guidelines or constructive recommendations are also lacking as to how to choose comparators in CEAs of screening. In order to gain some insights into ideal ICER estimation, this third part of the introduction reviews cancer screening CEAs that simulate thousands of screening strategies. All else equal, studies that simulate many strategies would be expected to have a higher chance to include all relevant strategies. This thesis conducted a rapid review of existing systematic reviews which record the number of strategies evaluated in the CEAs and chose three example analyses with over a thousand strategies in different types of cancer screening.

Cervical cancer screening

A Dutch CEA of cervical cancer screening published in two parallel papers by de Kok et al. (2012) and van Rosmalen et al. (2012) considered a total of 1,539 strategies. This CEA used the MISCAN cervix microsimulation model to simulate the strategies varied by screening modalities and screening schedules. Screening modalities include HPV testing and cytology, and different sequences or combinations of both modalities result in nine possible strategies. The screening initiates at the ages 25, 27, 30, or 32 with an interval of 3-10 years, comprising 3-10 times of screens. No screening is later than age

70. In total, 171 possible screening schedules are considered across 9 different combinations of primary and triage testing, yielding the 1,539 strategies in total.

Both publications explained how to apply incremental analysis for the calculation of ICERs in detail, including the definitions of simple dominance and extended dominance. Regarding the presentation of results, de Kok et al. (2012) did not provide numeric estimations of costs, effectiveness, and ICERs, while van Rosmalen et al. (2012) recorded the results only for the efficient frontier. In the publication of van Rosmalen et al. (2012), efficient screening strategies are detailed with ICERs, ranging from €3,701 to €122,508 per QALY.

The efficient frontier plotted on the cost-effectiveness plane can be found in both studies. Overall, the efficient frontier appears a relatively smooth curve as it is composed of 12 strategies. Cytology is only optimal when the cost-effectiveness threshold is below €7,000. The majority of efficient frontier is composed of HPV testing, which means HPV testing is dominant under most threshold scenarios. Along the efficient frontier, the number of screening rounds is observed to increase when their ICER increases, implying diminishing marginal returns of higher intensity screening.

The CEA separately identified and discussed the optimal strategy for the decision thresholds of €20,000 €30,000 and €50,000 per QALY, which generally have 3-5 rounds of screening in the lifetime. These optimal strategies are not located at the edge of the efficient frontier. Notably, the upper end of the efficient frontier is almost flat with a steep jump in ICERs from €57,514 to €122,508/QALY. This indicates that there may have been a shortage of higher intensity strategies with intervals of less than three years. This is unlikely to be a problem if the cost-effectiveness threshold is €50,000/QALY or less, but it could indicate insufficient strategy exploration had the relevant threshold been higher.

Breast cancer screening

A Spanish CEA of Vilaprinyo et al. (2014) applied the probabilistic model developed by Lee and Zelen (2008) to examine the cost-effectiveness of breast cancer screening. This mathematical model consists of a set of equations to estimate the cumulative probability of death for a specific cohort after years of follow-up. This model only has four health

states, including disease-free, preclinical, clinical, and death. This CEA considers alternative risk subgroups with subgroup-specific strategies, which greatly increases the potential number of strategies to consider. In particular, the choice of the analysts in this case to consider combinations of alternative strategies for alternative subgroups in a single analysis than considering multiple separate analyses where the stratification is considered ex-ante results in a large number of strategies to consider.

This CEA evaluated a total of 2,625 screening strategies. Most of the screening strategies are risk-based, combining different screening schedules for the separate risk groups. Only 24 are uniform screening strategies applied to all groups. A total of four risk groups are identified. Screening starts at age 40, 45 or 50 and ends at 69 or 74. The screening interval can be every one, two, three, or five years. The group with higher risk has higher-intensity screening. There is no variation in the use of screening modality, which is mammography for all.

This study obtained the efficient frontier and presented ICERs that accord correctly with the conventional interpretation. In addition to the cost-effectiveness plane, the numeric cost-effectiveness outcomes are also reported for the strategies on the efficient frontier. This CEA has an efficient frontier that appears smoothly curved consisting of 21 strategies. The ICERs range from €1,000 to €192,000 per QALY. Strategies with ICERs below €10,000 have screening every five years and strategies with ICERs above €58,000 have annual screening for all the risk groups. The strategies with ICERs between €10,000 and €58,000 have a combination of five-yearly and annual screening for different risk groups. With the exception of the most effective strategy overall of annual screening for all groups the policies offering risk-group specific strategies were more efficient than the policies offering a common strategy for all.

Bowel cancer screening

Whyte et al. (2022) used the University of Sheffield School of Health and Related Research (ScHARR) Bowel Cancer Screening Model (Whyte et al., 2011; Whyte et al., 2012) to simulate a large range of bowel cancer screening strategies. This CEA calculated net monetary benefit (NMB) to identify the optimal screening strategy under a constrained endoscopy capacity. Accordingly, this CEA did not report ICERs and appraise the efficient frontier.

This CEA considered a total of 60,221 screening strategies, which is an unusually large number of strategies to simulate. This CEA evaluated different possibilities of FIT cut-offs which is a substantial factor here as it amplifies the number of strategies considered. Screening is implemented with FIT thresholds between 20 and 180 micrograms of haemoglobin per gram of faeces. The screening starts at ages between 50 and 60 years old with a fixed screening interval of 1-6 years, having three or more than 3 rounds of screening and the FIT threshold is held constant within each strategy.

This CEA found numerous screening strategies have similar cost-effectiveness estimates under an annual colonoscopy capacity of 49,500-50,000 (Whyte et al., 2022). The characteristic of the 20 most cost-effective strategies with a restricted colonoscopy capacity of 50,000 was summarised by this CEA for policy recommendations. The ideal strategy is biennial screening starting at age 50-53 years old have 7-8 screening episodes with a FIT cut-off threshold of 127-166.

This analysis only reported the most cost-effective screening strategy is the most intensive screening strategy considered, which is annual screening with FIT 20 ages 50-74), for the scenario in the absence of any capacity constraint. With limited information, it is difficult for readers to understand if there is any chance to find other better or replaceable strategy, especially if similar outcomes are observed for several strategies with a colonoscopy capacity of 50,000.

Summary

Some commonalities among these three CEAs can be summarised. They have a powerful model capacity that can simulate an extensive range of strategies. This is not achievable for many conventional models, especially those coded in Excel. Unfortunately, these three models are not open-source and would be challenging to replicate. Without the ability to access the models and consider additional strategies, we cannot determine if the CEAs omitted relevant strategies that would lie beyond the efficient frontiers in each case.

All three CEAs only report partial results, as might be expected given the large number of strategies considered in each case. The CEAs of cervical cancer and breast cancer reported numeric results only for strategies on the efficient frontier. The CEA of bowel

cancer screening which identified optimal strategies by NMB only reported results for the 20 most cost-effective strategies. Notably, Vilaprinyo et al. (2014) presents the results of all the strategies on the cost-effectiveness plane. Although detailed numeric outcomes are still unavailable for these strategies, the cost-effectiveness plane displays the distribution of strategies in terms of cost and effectiveness estimates.

Each of the three CEAs defines the screening schedules in a systematic fashion which separately clarifies the interesting ranges of screening starting age, stop age and interval. The screening comparators are captured by permutation and combination of the factors associated with the screening policies. The combination is used to calculate the possible strategies with variations in factors that are independent of each other, such as starting ages, stop ages, and intervals. For example, de Kok et al. (2012) and van Rosmalen et al. (2012) multiply 9 modalities by 171 screening schedules, resulting in 1,539 possible strategies. The permutation is applied to the scenario when considering different strategies for risk groups. For example, in the CEA of Vilaprinyo et al. (2014), the higher risk group can only have the strategy with an earlier starting age, later stop age, and shorter interval than the lower risk group.

Regarding the scope of the strategies considered in each CEA, the studies did not specify how to choose the screening schedules and modalities simulated despite a large range of strategies. This might be a limitation of these CEAs. Furthermore, these CEA include their own research interest in the feature of strategies. Vilaprinyo et al. (2014) include the subgroup analyses with a single screening modality. de Kok et al. (2012) and van Rosmalen et al. (2012) have a variation in screening modalities but only applied to one population. Whyte et al. (2022) have the main variation in the characteristics of screening modalities, FIT cut-offs. Considering the variation in one feature has resulted in a large number of strategies, it will be more challenging for a study to consider all the possible variations.

Regarding the simulation results, Vilaprinyo et al. (2014), de Kok et al. (2012), and van Rosmalen et al. (2012) have efficient frontiers that appear to be smoothly curved and a range of ICERs that covers the range of decision thresholds. Although the efficient frontier is not visually presented on the cost-effectiveness plane in the CEA of Whyte et al. (2022), considering the very many screening strategies simulated, the gap between the

strategies is likely to be very small. A smooth efficient frontier is accordingly also expected from the results of Whyte et al. (2022).

All three CEAs are applied research, which cannot really explain the good practice of strategy selection and comparison in detail. These CEAs chose the studied strategies before the simulation started and they did not discuss the possibilities of strategy omission in their CEAs. None of the three CEAs explained the characteristics of strategies on the efficient frontier and none of them discussed the parameter that influences the ICERs of the efficient strategies. This can be interpreted as lacking exploration of the impact of the characteristics of strategies, including screening schedule and modalities, and other model parameters, on their cost-effectiveness.

The research gap

The following material describes the research gap addressed by this thesis. It draws on the literature introduced above to explain what issues have not yet been addressed cogently within the CEA literature on cancer screening.

Many textbooks and guidelines teach how to conduct a CEA in general but very few discuss how to conduct CEA in the specific context of cancer screening. For example, "cancer screening" is only mentioned twice in the prominent textbook by Drummond et al. (2015). Although the textbook by Gold et al. (1996) used some examples of cancer screening to illustrate the methodology of CEA, it did give detailed practical guidance of attempting to implement what is recognised as important in theory. For example, Gold et al. (1996) suggested that CEA use the principle of including all the frequencies that are feasible for screening programmes. This recommendation, however, does not contain details to guide an applied modeller regarding what range of strategies to consider.

The existing checklists for economic evaluation do not adequately reflect relevant methodological concerns regarding strategy choice and comparison in screening CEAs. Most checklists emphasise the disclosure of method (transparency), instead of the research design. For example, a recognised methodological deficiency in the CEAs, the confusion in ACER and ICER (Drummond & Sculpher, 2005), is not properly examined in the existing checklists. In addition, the choice of screening strategies is not examined

in existing checklists. This is an essential methodological issue as strategy omission can lead to a biased ICER estimate.

As described in Part 2 of the introduction, there is no open-source teaching model or material designed for screening CEAs. Existing simplified models are designed for other purposes. An example is the Sick-Sicker model which is applied as a tutorial for building a Markov model in R while it does not consider the intervention of screening (Alarid-Escudero et al., 2022). Most screening CEAs models focus on the applied research rather than the simulation methods or tutorials. They are too complex and not suitable to illustrate the methodological concerns in the screening context, such as the comparator omission.

The transparency of CEA models is another factor hindering the improvement of methodology applied in cancer screening CEAs. Many CEAs are published without complete model computation descriptions, preventing others from learning from them or extending their research. The illustration of issues in simulation methods typically require the application of a full model, which means they can only be investigated by groups that hold such models. The reluctance to share complete model code leads to a limited distribution of such models and a barrier for other researchers seeking to improve simulation methodology.

There is a research gap regarding the inclusion of an inadequate range of strategies in cancer screening CEAs. Strategy omission is a recognised and fundamental methodological problem in the CEAs of screening but is not clearly documented in guidelines. This research gap is partly due to the lack of clear and detailed methods recommendations regarding comparator choice in the CEAs of screening from existing textbooks, guidelines, checklists, teaching models and tutorials. Also, the absence of an open-source screening model designed for illustrative purposes is another reason why problems in screening methods may persist in the cancer screening CEA literature.

Ensuring the reliability of CEAs is critical for decision-makers. It is important that decision-makers are presented with sound evidence-based policy advice that is free from systematic methodological biases. Biased CEAs cannot off sound support to decision-making and eventually leads to a sub-optimal screening policy. Not only does persisting

with biased modelling methods risk sub-optimal screening policy but it also results in research waste as research time and resources are committed to studies that do not generate the right answers.

Aims of this thesis

The purpose of this thesis is to increase the reliability of health economic evidence in the cancer screening literature, in terms of the identification of the optimal strategy. Overall, the objectives are (1) to increase our understanding of the cost-effectiveness of particular interventions, in terms of cancer screening; (2) to provide a pedagogical model of screening; and, (3) to improve the understanding of the impact of strategy choices on the identification of the optimal strategy in the screening CEAs. The thesis posits that the predominant objective of the majority of CEAs in screening is to facilitate decision-making for screening policies by identifying the most cost-effective strategies.

Study outlines

Study 1 reviewed the choices of strategies and their comparison in CEAs of CRC screening. This thesis chose CRC as an example because this methodological issue has not been examined in this specific context previously. The exploration in the CRC context summarised the issues of ICER calculation and the inadequate strategies inclusion and demonstrates how it might lead to strategy omission. Study 2 then developed a pedagogical model for screening to understand the impacts of key parameters on the cost-effectiveness of screening strategies and the efficient frontiers. Finally, Study 3 presents an application of this pedagogical model to explain how to avoid strategy omission from a perspective of the characteristics of screening schedules, the range of ICERs, or the convexity of efficient frontier.

This thesis also touched on the study in Appendix 1. As I am not the first author of this study, I did not include the study details in the main thesis. Appendix 1 is a review summarising the characteristics of strategies considered in the CEAs of CRC screening in Europe. The findings from this review are also compared and discussed in conjunction with the results of the first study in this thesis.

Study 1: a review of the CEAs of CRC screening

The first study addresses the appropriateness of cost-effective methodologies applied in the CEAs of CRC screening, including the strategy choice, their comparisons and the ICER calculation. This thesis developed a novel appraisal tool as the existing checklists cannot adequately address these methodological issues. This thesis reviewed the published CEAs on CRC screening cost-effectiveness and appraised their quality of the identification of cost-effective strategies. To objectively understand the need for CEA pedagogical research, this study will calculate the proportion of the CEAs that fail to demonstrate an adequate range of screening strategies and fail to estimate corresponding ICERs for the decision-making context. This review will focus on CRC because this issue has not yet been examined in this cancer in the existing literature.

Study 2: a pedagogical model for CEA of cancer screening

This study develops a simple, open-source modelling framework to illustrate the effect of the choice of comparator strategies and model parameters on the simulation of efficient frontiers. This study adopts the R programme due to its rapid growth of applications in CEA modelling. This study provides the basic core model code and supporting documentation. This study uses the model to illustrate how the position of the efficient frontier changes as key parameters vary, including the screening test performance, the disease history, and the incidence of cancer. This study employs comparative statics to present the efficient frontiers on the cost-effectiveness plane.

Study 3: an ideal efficient frontier of cancer screening

This study applies the framework proposed from the second study to support the development of a clear understanding of the appropriate ICER generation and shows how the problems of comparator omission can be avoided. This study observes the distribution of strategies with constant characteristics of screening schedules, including fixed screening start age, stop age, or the screening interval. Furthermore, this study explains the importance of the policy-relevant portion of the efficient frontier which theoretically influences the final decision. Last, this study extends the results from the second study to observe the impact of parameter changes on the convexity of efficient frontier. This study provides a simple guide for CEA analysts to self-examine if they omit critical comparators by observing their cost-effectiveness results.

2. STUDY 1 A REVIEW OF THE CEAS OF CRC SCREENING

This chapter describes the first study of this thesis, which is a systematic review examining the appropriateness of cost-effective methodologies applied in the CEAs of screening, specifically regarding cost-effective strategy identification. The review utilized a self-developed evaluation form to assess the strategy choices and their comparisons for calculating ICERs in the CEAs. It focuses on CRC since the current literature lacks examination of this aspect in the context of this cancer.

Methods

This section specifies how this thesis conducted the systematic review of CEAs of CRC screening, applied the self-developed quality assessment tool for the evaluation of comparator choices, and established the hypotheses for statistical analysis. For the systematic review, this section separately describes the data sources, study selection, and data extraction. For the quality evaluation, this section explains each item in the assessment tool and gives the reason why the item was added to the tool. For statistical analysis, a series of hypotheses are given to examine the association between features of the CEAs and their scores in the quality assessment.

Systematic review

Data sources

PubMed, Web of Science, and Scopus databases were searched up from their inception to 4th September 2020. Relevant articles identified during the search are also included in this review. This review also undertook citation tracking, retrieving CEAs cited by existing reviews (Hanly et al., 2012; Jeong & Cairns, 2013; Kriza et al., 2013; Lansdorp-Vogelaar et al., 2011; Mendivil et al., 2019; Patel & Kilgore, 2015; Pignone et al., 2002; Ran et al., 2019; Skally et al., 2013). The research protocol was registered with the international prospective register of systematic reviews (PROSPERO, registration number: CRD42020147532).

The key concepts of the research question for this systematic review are summarized using the PICOS framework in Table 2-1, which includes Participant (P), Intervention (I), Comparator (C), Outcome (O), and Study design (S). The corresponding search

strings are developed and detailed in Table 2-2, Table 2-3, and Table 2-4, respectively for Web of Science, Scopus, and PubMed databases. The string was developed with a series of categories for the search, including the studied disease (CRC), intervention (screening), study type (CEA), and study outcomes (cost per QALY or LY gained).

Study inclusion and exclusion

We included applied, model-based CEAs of average-risk, population-based CRC screening programmes. This review considered any screening modality for the secondary prevention of CRC, including stool and image-based tests. Stool-based modalities include the FOBT, FIT, and faecal DNA (fDNA) test. Image-based modalities include sigmoidoscopy, colonoscopy, CTC, and barium enema (BE).

While the exact definition of average-risk may vary between applied studies, this review typically excludes those with symptomatic disease and those of known risk groups such as individuals with Lynch Syndrome or a close family history of disease. We excluded methodological studies and CEAs assessing hypothetical screening modalities not based on currently observed estimates of test performance. Similarly, we excluded analyses of interventions promoting screening adherence. We excluded CEAs that reported outcomes other than cost per QALY or LY gained, such as cost per case detected, colonoscopies per LY gained, or NMB. We also excluded studies not written in English.

Two reviewers were responsible for retrieving and examining the data. If needed, a third reviewer helped achieve consensus. All searches were conducted in the online databases. Selection and reviewing of the retrieved records were conducted in EndNote X8 and Microsoft Excel 2019.

Table 2-1 Research question of the systematic review using the PICOS framework

PICOS	**Search strategy**
Participant (P)	The general population
Intervention (I)	Any clinical-based colorectal cancer screening modalities, including faecal occult blood test (FOBT), faecal immunochemical test (FIT), faecal DNA (fDNA), colonoscopy, sigmoidoscopy, computed tomography colonography (CTC), and barium enema (BE).
Comparator (C)	
Outcome (O)	Cost per quality-adjusted life-year (QALY) and cost per life-year (LY).
Study design (S)	Cost-effectiveness analysis (CEA)

The search string is developed based on the concepts outlined in this PICOS framework. The interventions and comparators are combined because this review is not focused on the incremental benefit of a particular intervention compared to others. Instead, this review aims to assess whether CEAs include adequate screening strategies for the purpose of identifying the optimal strategy, for which incremental comparison should be made between the strategies on the frontier.

Table 2-2 The full search strategy (Web of Science)

Search concept	**Search string**
Colorectal	TS=(colon* OR rectal OR colorectal)
Cancer	TS=(cancer* OR tumour* OR tumor* OR neoplasm* OR carcinoma* OR adenoma* OR polyp*)
Screening	TS=(screen* OR colonoscop* OR sigmoidoscop* OR FSIG OR computer tomography colonography OR CTC OR faecal occult blood OR fecal occult blood OR *FOBT OR faecal immunochemical test OR fecal immunochemical test OR FIT OR barium enema OR stool DNA testing)
Cost-effectiveness filter	TS=(cost effect* OR CEA OR cost utility OR CUA OR health technology assessment OR HTA OR health economics OR costs OR cost benef* OR cost anal*)
Cost-effectiveness outcomes	TS=(quality adjusted OR QALY* OR disability adjusted OR DALY* OR LYG OR life-year* OR life expectancy)
1 and 2 and 3 and 4	
Limit to English language and original articles	

CEA, cost-effectiveness analysis; CTC, computerised tomography colonography; CUA, cost-utility analysis; DALY, disability adjusted life-years; FIT, immunochemical faecal occult blood test; FOBT, faecal occult blood test; FSIG, flexible sigmoidoscopy; HTA, health technology assessment; LYG, life-years gained; QALY, quality-adjusted life-years.

Table 2-3 The full search strategy (Scopus)

Search concept	Search string
Colorectal	TITLE-ABS(colon* OR rectal OR colorectal)
Cancer	TITLE-ABS(cancer* OR tumour* OR tumor* OR neoplasm* OR carcinoma* OR adenoma* OR polyp*)
Screening	TITLE-ABS (screen* OR colonoscop* OR sigmoidoscop* OR FSIG OR "computer tomography colonography" OR CTC OR "faecal occult blood" OR "fecal occult blood" OR *FOBT OR "faecal immunochemical test" OR "fecal immunochemical test" OR FIT OR "barium enema" OR "stool DNA testing")
Cost-effectiveness filter	TITLE-ABS ("cost effect*" OR CEA OR "cost utility" OR CUA OR "health technology assessment" OR HTA OR "health economics" OR costs OR "cost benef*" OR "cost anal*")
Cost-effectiveness outcomes	TITLE-ABS ("quality adjusted" OR QALY* OR "disability adjusted" OR DALY* OR LYG OR "life year*" OR "life expectancy")
1 and 2 and 3 and 4	
Limit to English language and original articles	

CEA, cost-effectiveness analysis; CTC, computerised tomography colonography; CUA, cost-utility analysis; DALY, disability adjusted life-years; FIT, immunochemical faecal occult blood test; FOBT, faecal occult blood test; FSIG, flexible sigmoidoscopy; HTA, health technology assessment; LYG, life-years gained; QALY, quality-adjusted life-years.

Table 2-4 The full search strategy (PubMed)

Search concept	Search string
Colorectal	colon*[tiab] OR rectal[tiab] OR colorectal[tiab]
Cancer	cancer*[tiab] OR tumour*[tiab] OR tumor*[tiab] OR neoplasm*[tiab] OR carcinoma*[tiab] OR adenoma*[tiab] OR polyp*[tiab]
Screening	screen*[tiab] OR colonoscop*[tiab] OR sigmoidoscop*[tiab] OR FSIG[tiab] OR computer tomography colonography[tiab] OR CTC[tiab] OR faecal occult blood[tiab] OR fecal occult blood[tiab] OR *FOBT[tiab] OR faecal immunochemical test[tiab] OR fecal immunochemical test[tiab] OR FIT[tiab] OR barium enema[tiab] OR stool DNA testing[tiab]
Cost-effectiveness filter	cost-effect*[tiab] OR CEA[tiab] OR cost-utility[tiab] OR CUA[tiab] OR health technology assessment[tiab] OR HTA[tiab] OR health economics[tiab] OR costs[tiab] OR cost-benef*[tiab] OR cost anal*[tiab]
Cost-effectiveness outcomes	quality-adjusted[tiab] OR QALY*[tiab] OR disability-adjusted[tiab] OR DALY*[tiab] OR LYG[tiab] OR life-year*[tiab] OR life expectancy[tiab]
1 and 2 and 3 and 4 and 5	
Limit to English language and original articles	

CEA, cost-effectiveness analysis; CTC, computerised tomography colonography; CUA, cost-utility analysis; DALY, disability adjusted life-years; FIT, immunochemical faecal occult blood test; FOBT, faecal occult blood test; FSIG, flexible sigmoidoscopy; HTA, health technology assessment; LYG, life-years gained; QALY, quality-adjusted life-years.

Data extraction

After identifying relevant CEAs, this review extracted details including authors, publication year, country setting, cost-effectiveness decision threshold, and features of the simulated strategies, including their test modality, screening start and stop ages, and the intervals between screens. This review extracted the principal health economic outcomes for each intervention (including health benefits, costs, and reported ICERs).

Quality assessment form

This review developed and applied our own quality assessment form by selecting and adapting items from CHEERS and Drummond checklists and adding additional questions (Drummond et al., 2015; Husereau et al., 2013). Such adaptation was required as existing checklists do not offer questions of sufficient relevance to CEAs of screening (Table 2-5). Regarding relevant quality assessment in CHEERS and Drummond checklists, the first relates to the complete reporting of cost and effects estimates, which is taken from the item 19th of the CHEERS checklist. The more specific checklist items then relate to the selection and comparison of alternative strategies and are adopted from the 2nd and 8th items in the Drummond checklist. Further details of both checklists can be found in the introduction of this thesis.

A total of three relevant parts are considered, including the completeness of reporting, the comparison used for ICER calculation, and the inclusion of strategies. If a CEA did not disclose enough details of the estimated costs and effects, this review cannot access its ICER calculation. The issue of transparency of results reporting and the problem of ICER calculation are independent concerns, so this review separates them into two parts. Table 2-5 outlines the quality assessment form and the following material provides further explanation of the three relevant parts.

Table 2-5 Quality assessment form

Quality Assessment Form

Part 1: The completeness of reporting

1. Does the study report the outcome estimates (both cost and health effects) numerically?

2. Are the outcome estimates reported to sufficient significant figures to enable adequate replication of the reported ICERs?

3. Are the outcome estimates reported for all the strategies?

4. Does the study present the cost-effectiveness plane?

Part 2: The ICER calculation

5. Are ICERs from incremental comparisons, rather than comparisons to no-screening?

6. Are ICERs only produced by the comparisons among efficient comparators?

Part 3: The adequacy of strategies inclusion

7. Does the study include both stool and image-based modalities?

8. Does the study only include stool-based modalities?

9. Does the study only include image-based modalities?

10. Are consistent screening schedules applied to the same type of modalities?

11. Is a no-screening option included (or are results reported relative to a no screening baseline)?

12. Are there any additional strategies mentioned within the study but not included in the base-case analysis?

13. Does the range of efficient frontier encompass the threshold?

14. Are the ICERs exclusively above the threshold?

15. Are the ICERs exclusively below threshold?

16. Is there no ICER estimate applicable (less than two strategies on efficient frontier)?

ICER, incremental cost-effectiveness ratio. This quality assessment form is designed for the evaluation of cost-effectiveness analysis of screening in terms of strategy inclusion and comparison for ICER calculation.

Part 1: The completeness of reporting

Part one of the assessment addresses completeness of reporting. It assesses if results are reported for all strategies. It also assesses if costs and effects were reported with sufficient precision to replicate the reported ICERs. Although it is typically not possible to replicate all reported ICERs exactly due to numerical rounding, it often remains clear what

comparisons have been made. Thus, this review only considered outcomes to be reported with insufficient precision where the reporting was insufficiently precise to permit identification of what comparisons between interventions were made. In addition, this review considers visual presentation of the outcomes of all the strategies on the cost-effectiveness plane may be helpful for the identification of efficient frontier and its corresponding ICERs. Accordingly, the assessment records if a cost-effectiveness plane was presented in the CEA.

Part 2: The ICER calculation

Part two assesses if correct comparisons have been made when calculating ICERs. For example, to estimate the ICER of strategy A relative to a comparator of strategy B the following formula would be used (ICER = $(COST_{strategy\ A} - COST_{strategy\ B})/(EFFECT_{strategy\ A} - EFFECT_{strategy\ B})$). There are two primary ways inappropriate comparisons are made. One is simply to compare all strategies relative to the no-screening scenario $(COST_{strategy\ A} - COST_{no\text{-}screening})/(EFFECT_{strategy\ A} - EFFECT_{no\text{-}screening})$. Such a comparison may be legitimate for the first ICER but will not in general be correct. Secondly to compare to a strategy that is not adjacent on the efficient frontier. A third form of inappropriate comparison that is typically less commonly observed is by simply assessing the ratio of gross costs to effects of a strategy without making any incremental comparison, to other strategies or, importantly a no-screening scenario. With the assessment, this review collectively considers the first and third examples as non-incremental comparisons. This review considers the second example to represent an incremental comparison with an incorrect comparator.

Part 3: The adequacy of strategies inclusion

Part three of the assessment addresses the completeness of the strategies compared. This review is not aware of any existing criteria for objectively appraising if sufficient strategies have been included in CEAs. This review developed two general principles to appraise this issue as objectively as possible, though this review recognises that they may not be applicable to all cases given the variety of analyses. These two principles separately examined the range of alternatives and cost-effectiveness results in each CEA.

The first principle relates to the variation of the strategies considered in the CEAs. This review assessed variation in testing modalities compared and screening schedules

simulated. Stool and image-based modalities are the two broad classes of interventions used in CRC screening, though it is also possible to have strategies that combine both classes of modality. An example of such a combination is Knudsen et al. (2010) which featured sigmoidoscopy every five years in conjunction with annual FOBT. This review considered an analysis to have included insufficient strategies if it only included either stool or image-based modalities.

Regarding screening schedules, to avoid varying multiple strategy characteristics simultaneously, this review expected the same age range and intervals to be considered in alternative modalities within either the class of stool or image-based screening. For example, if an analysis considered annual and biennial screening FOBT and it assessed FIT, then this review assumed it should assess those same screening intervals in FIT but not necessarily in an image-based modality such as colonoscopy. Similarly, if CTC was considered over a screening age range of 50 to 70 and the analysis also considered colonoscopy, this review assumed it should apply the same age range to that colonoscopy, but not necessarily to any stool-based strategies considered. Analyses that applied different age ranges and intervals within common classes of modality were recorded as applying inconsistent screening schedules.

Existing guidelines recommend economic evaluations consider a do-nothing scenario (Drummond et al., 2015). Accordingly, this review assessed if CEAs reported estimates for a no-screening scenario or, equivalently, reported results relative to no-screening. This review also assessed if studies compared all the simulated screening strategies in the base-case analysis as opposed to only simulating certain strategies in sensitivity analyses or other sub-analyses.

The second principle of assessing strategy completeness relates to the variation in cost-effectiveness estimates made by the studies. This review considered the range of strategies simulated and the resulting efficient frontier. Studies that consider a large variety of screening intensities typically feature efficient frontiers made up of numerous strategies and have ICERs that rise from below the cost-effectiveness threshold and extend above it. Such frontiers offer some reassurance that the analysis includes sufficient margins of low and high intensity strategies to ensure that those strategies with ICERs close to cost-effectiveness threshold truly are those most likely to be efficient.

This review expects that the CEAs' reported ICERs encompass the range of country-specific thresholds because cost and effectiveness estimates can be varied by adjusting the screening intensity. For example, in the UK context which has a threshold range of £20,000-£30,000/QALY, if a CEA has an efficient frontier with ICERs that range only between £5,000-£10,000/QALY, then it is likely that higher-intensity strategies with ICERs closer to the threshold have been omitted. Conversely, if all the estimated ICERs were above £50,000/QALY, then it is quite possible that lower intensity, more cost-effective strategies have been omitted. Accordingly, this review assessed if the estimated efficient frontiers had ICERs that encompassed the threshold or lay exclusively above or below it. This assessment used the relevant threshold mentioned within the study. If no threshold was mentioned, this review used the most frequent threshold mentioned in other CEAs from the same jurisdiction.

Statistical analysis

Though not part of our original analysis plan as documented in the submitted study protocol, this thesis also assessed if certain study features were related to quality performance according to the assessment form. The first hypothesis of this study was that older CEAs would be of lower quality. Another hypothesis was the inclusion of a cost-effectiveness plane would be associated with better quality scores. This review was also sought to explore regional differences of quality performance. This study tested the hypothesis that there was a difference in the quality scores across continents, including North America, Europe, Australasia, and Asia. The hypotheses were investigated with Chi-squared test or Fisher's exact test at the 5% level of significance in Rstudio Version 1.1.463.

Results

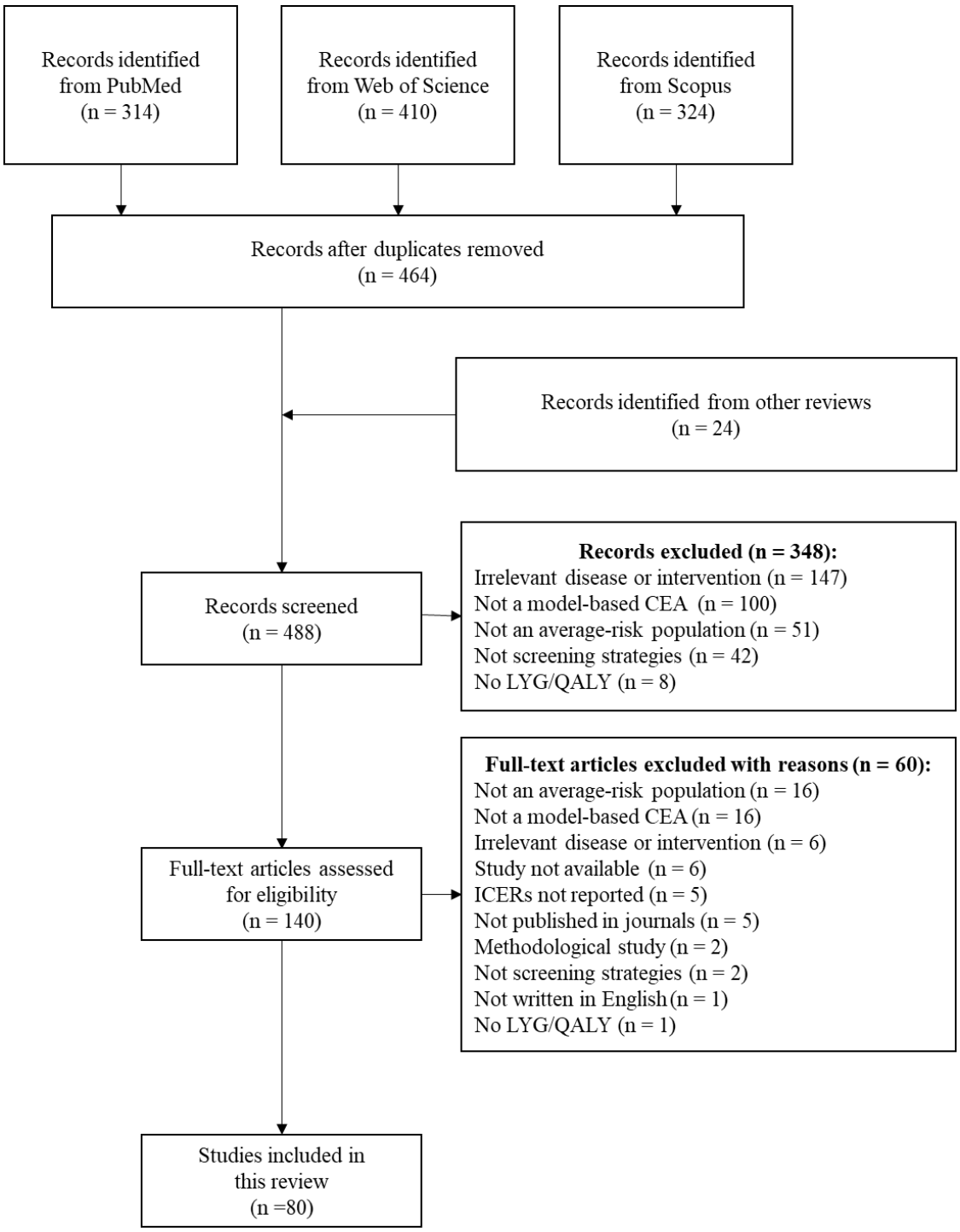
This section described the results of the systematic review of CRC screening, including the search results from the databases, background characteristics of the included CEAs, the results of the quality assessment, and the test outcomes of associative hypotheses. This study summarised the CEAs' publication year, country setting, the number of strategies included, and the frequency of screening modalities applied. This study

detailed the outcomes for each item in the self-developed tool and examined the association between features of the CEAs and their scores in the quality assessment.

Search results

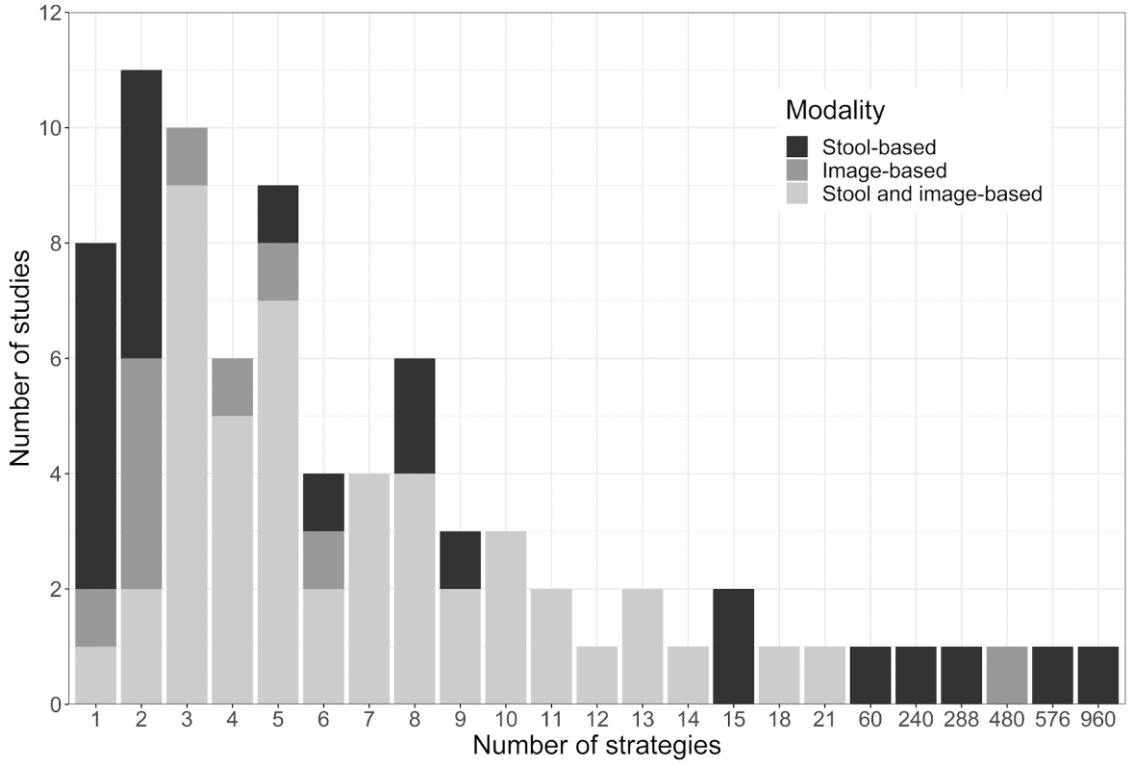
The search generated 464 unique studies. Searching existing reviews identified a further 24. There were 348 studies excluded at title and abstract screening. The most common reasons for exclusion were irrelevant disease or intervention (n = 147) and not being a model-based CEA (n = 100). Full-text articles were assessed for the remaining 140 studies. The reasons for excluding a further 60 studies are recorded in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram in Figure 2-1. A total of 80 CEAs of CRC screening were reviewed. Notably, the study by Peterse et al. (2021) has been accessible online since August 2020. Although its journal publication date is beyond our predefined inclusion criteria, this review intentionally incorporated this CEA, in accordance with the registered protocol that specified the inclusion of any CEAs available before September 4th, 2020.

Figure 2-1 The PRISMA flow diagram



PRISMA, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Figure 2-2 Frequency of study by number of strategies simulated



This histogram summarises the number of studies categorised by the number of screening strategies simulated in the base-case analysis.

Descriptive analysis

Of the 80 CEAs, only 9 CEAs (11%) were published from 1991 to 2000; 26 (33%) were published from 2001 to 2010; and 45 (56%) have been published from 2011 to 2020. The most common country settings were the US (n = 28, 35%), the Netherlands (n = 8, 10%), and Australia (n = 6, 8%). Most CEAs (n = 64, 80%) had ≤ 10 strategies (excluding no-screening), and only 5 (6%) included more than 100 strategies (Figure 2-2). Notably, these five studies only assessed stool-based modalities. The most commonly assessed modalities were colonoscopy (n = 51, 64%), FIT (n = 49, 61%), and FOBT (n = 47, 59%). The least common modalities assessed were BE (n = 6, 8%), fDNA (n = 9, 11%), and CTC (n = 15, 19%). The proportion of CEAs analysing stool-based modalities in each continent ranged from 79% (n = 27) in North America to 100% (n = 7) in Australasia. In contrast, the proportion of CEAs analysing image-based modalities ranged wide, from 29% (n = 2) in Australasia, 50% (n = 14) in Europe, to 91% (n = 31) in both North America and Asia.

Table 2-6 The quality assessment of the cost-effectiveness analysis of colorectal cancer screening (N = 80)

	Yes		Insufficient information[†]	
	N	%[††]	N	%
Completeness of reporting				
Report both cost and effect estimates	75	95%		
Sufficient significant figures[†††]	70	93%	2	3%
All the strategies[†††]	67	89%		
Cost-effectiveness plane	47	59%		
ICER calculation				
Incremental comparisons[††††]	39	49%	3	4%
Correct comparators	49	61%	4	5%
Adequacy of strategy inclusion				
Both stool and image-based modalities included	47	59%		
Only stool-based modalities	23	29%		
Only image-based modalities	10	13%		
Consistent screening schedules	47	59%		
No-screening option included	75	94%		
No additional strategies in sensitivity analysis[†††]	72	94%		
The range of ICERs encompasses the threshold	18	23%	7	9%
ICERs exclusively above threshold	7	9%		
ICERs exclusively below threshold	41	51%		
No ICER reported	7	9%		

ICERs, incremental cost-effectiveness ratios.
† The quality assessment was constrained by the limited information reported in the CEAs.
†† The percentage was calculated with the number of CEAs with sufficient information.
††† The percentage was calculated by the number of studies reporting the cost/effects or having sensitivity analysis.
†††† Non-incremental comparisons include cost-effectiveness ratio without comparing to any intervention and ratio compared to no-screening.
This evaluation only focuses on the quality of strategy inclusion and comparison for ICER calculation. The details of the assessment items can be found in the Methods section of this study.

Quality assessment

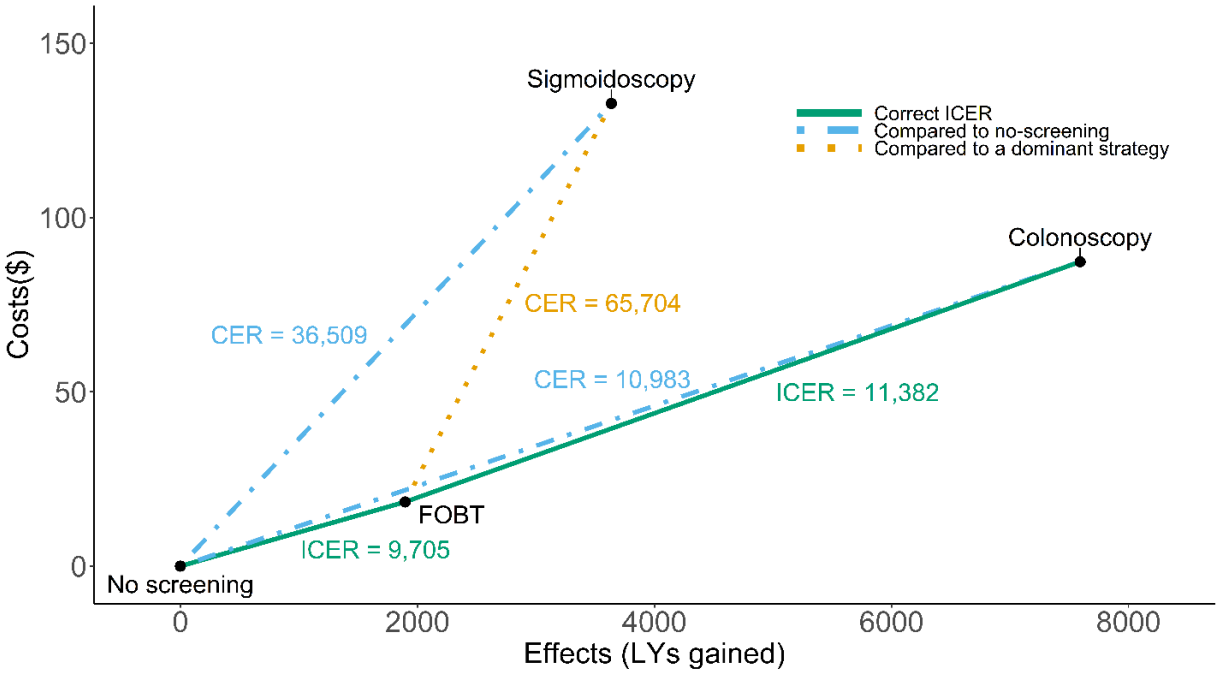
Table 2-6 summarises our quality assessment results. Overall, there was good performance regarding completeness of reporting. Most (n = 75, 94%) reported both cost and effects estimates numerically, of those, most reported costs and effects to sufficient accuracy to replicate ICERs (n = 70, 93%). While the proportion of studies reporting estimates for all simulated strategies numerically was 89% (n = 67), only 59% (n = 47) featured a cost-effectiveness plane.

Regarding the calculation of ICERs, many studies did not perform well. Of all 80 CEAs, only 46% (n = 37) did not include any ambiguities in the reported ICERs and 5% (n= 4) CEAs that could not be assessed because they did not report estimates numerically. The most common misinterpretations were the reporting of ratios compared to no-screening as ICERs (n = 38, 48%). For example, Tappenden et al. (2007) reported all the results based on the comparisons to the no-screening scenario. Estimating ICERs on the basis of comparisons to dominated comparators were found in one-third of CEAs (n = 27, 34%). An example is that of Telford et al. (2010) which ICERs were simply compared to the previous less effective and costly strategy without ruling out dominated strategies, although they reported correct ICERs in their appendix. Of the 80 CEAs, 22 (28%) included both types of ambiguous comparisons as ICERs estimated relative to no-screening (non-incremental comparisons) and relative to dominated strategies (incorrect comparators).

A notable feature of the literature is the prevalence of tables in which alternative strategies are listed as both row and column headings and the CERs of each strategy is cross-tabulated against each other. Figure 2-3 reproduces the earliest example found which is from Sonnenberg *et al.* (2000). The pair-wise comparisons in such tables necessarily include correctly interpreted ICERs, but also feature other ratios based on comparisons between strategies that are not adjacent on the efficient frontier and so are inconsistent with the definition of an ICER. This review found 16 (20%) studies reporting such tables.

Figure 2-3 Incremental cost-effectiveness ratios reported in Sonnenberg et al. (2000)

A.Cost-effectiveness plane



B. Inappropriate reporting format

	No screening	FOBT	Sigmoidoscopy	Colonoscopy
No screening	0	9,705	36,509	10,983
FOBT	-	0	65,704	11,382
Sigmoidoscopy	-	-	0	Dominates
Colonoscopy	-	-	-	0

C. Ideal reporting format

	Total Cost ($)	Total Effectiveness (LYs)	Incr. Cost ($)	Incr. Effectivesness (LYs)	ICER ($, Incr. Cost/Incr. LY)
No screening	136,452,922	0	-	-	-
FOBT	154,853,577	1,896	18,400,655	1,896	9,705
Sigmoidoscopy	269,214,301	3,636	-	-	Dominated
Colonoscopy	223,780,829	7,952	68,927,252	6,056	11,382

CER, cost-effectiveness ratio; FOBT, faecal occult blood test; ICER, incremental cost-effectiveness ratio; Incr., incremental; LY, life-year. The labelled numbers are ICERs expressed in dollars ($) per life-year gained. Table B is extracted from Sonnenberg et al.(2000) and this study further plots the results on the cost-effectiveness plane to illustrate the correct and incorrect ICERs.

Regarding the omission of relevant strategies, many CEAs failed at least one item on our assessment for including adequate strategies. Only 59% (n = 47) included both stool and image-based modalities. Compared to studies exclusively assessing image-based modalities (n = 10, 13%), more than twice as many CEAs exclusively assessed stool-based modalities (n = 23, 29%).

Only 59% (n = 47) included consistent screening schedules across the same type of modality. Frequently observed examples were analyses considering sigmoidoscopy every 5 and 10 years but only evaluating colonoscopy every 10 years. In such cases there appears scope to assess the cost-effectiveness of colonoscopy every 5 years too. Failing to consider a 5-year interval represents an example of omitting a potentially relevant strategy, especially if the ICER of colonoscopy every 10 years falls within the cost-effectiveness threshold, indicating that there may still be scope for further intensification of screening to remain cost-effective. Another common example was the simulation of annual FIT and FOBT when fDNA was only modelled every three years. Although fDNA is more expensive than other stool-based modalities, if it is cost-effective at every three years, then it is worth examining if it remains cost-effective at higher frequencies. This review found 22 (28%) studies that simulated strategies that combine both stool and image-based modalities, particularly combining sigmoidoscopy with FIT or FOBT. Notably, these strategies were sometimes the most cost-effective. In the example of Naber et al. (2019), 10-yearly flexible sigmoidoscopy was combined with annual FOBT or annual FIT and they found the combination of FOBT was efficient in two models out of three.

Regarding the range of ICERs relative to the cost-effectiveness threshold, this review found only 40% (n=32) of studies reported a cost-effectiveness threshold, but thresholds could be inferred from the remaining studies from other analyses within the sample. Only 23% (n = 18) of studies appraised an efficient frontier that ranged from below to above the relevant cost-effectiveness threshold. Most (n = 41, 51%) had ICERs exclusively below the threshold, 7 (9%) had ICERs exclusively above the threshold. Regarding the number of strategies on the frontier, only 2 (3%) studies found ≥10 strategies on the frontier. Most identified two (n = 28, 35%) or three (n = 20, 25%) strategies on the frontier.

Table 2-7 The detailed results according to Part 1 of our quality assessment: the completeness of reporting

Authors (Year)	Country/ Region	No. strategies	Report threshold	Report numeric estimates	Report sufficient figures	Report all results	Cost-effectiveness plane
Areia et al. (2019)	Portugal	2	Y	Y	Y	Y	Y
Aronsson et al. (2017)	Sweden	4	Y	Y	Y	Y	N
Azad et al. (2020)	US	5	Y	Y	Y	Y	Y
Barouni et al. (2012)	Iran	3	Y	Y	Y	N	N
Barzi et al. (2017)	US	13	Y	Y	Y	Y	Y
Berchi et al. (2004)	France	2	N	Y	Y	Y	N
Coldman et al. (2017)	Canada	8	N	N	NA	NA	N
Coretti et al. (2020)	Italy	1	N	Y	Y	Y	Y
Dan et al. (2012)	Singapore	10	Y	Y	N	Y	Y
Dinh et al. (2013)	US	5	N	Y	Y	Y	Y
Fitch et al. (2015)	US	1	N	Y	N	Y	N
Flanagan et al. (2003)	Canada	2	N	N	NA	NA	N
Frazier et al. (2000)	Canada	21	N	Y	Y	Y	Y
Goede et al. (2013)	Netherlands	960	N	Y	Y	N	Y
Goede et al. (2017)	Canada	576	Y	Y	Y	N	Y
Greuter et al. (2016)	Netherlands	12	Y	Y	Y	Y	Y
Greuter et al. (2017)	Netherlands	5	Y	Y	Y	Y	Y
Gyrd-Hansen (1998)	Denmark	15	N	N	NA	NA	Y
Gyrd-Hansen et al. (1998)	Denmark	60	Y	Y	Y	N	Y
Hassan et al. (2008)	US	3	Y	Y	Y	Y	Y
Hassan et al. (2011)	France	9	Y	Y	Y	Y	Y
Heitman et al. (2010)	North America	10	N	Y	Y	Y	N
Heresbach et al. (2010)	France	3	Y	N	NA	NA	Y
Huang et al. (2014)	China	8	N	Y	Y	Y	N
Jahn et al. (2019)	Austria	3	Y	Y	Y	Y	Y
Khandker et al. (2000)	US	8	N	Y	Y	Y	N
Knudsen et al. (2010)	US	14	Y	Y	Y	Y	Y
Ladabaum and Mannalithara (2016)	US	4	Y	Y	Y	Y	Y

(Continued)

Authors (Year)	Country/ Region	No. strategies	Report threshold	Report numeric estimates	Report sufficient figures	Report all results	Cost-effectiveness plane
Ladabaum et al. (2004)	US	2	N	Y	Y	Y	N
Ladabaum et al. (2013)	US	8	Y	Y	Y	Y	Y
Ladabaum et al. (2014)	Germany	11	Y	Y	Y	Y	Y
Ladabaum et al. (2019)	US	7	N	Y	Y	Y	N
Lansdorp-Vogelaar et al. (2010)	US	18	Y	Y	Y	Y	Y
Lee et al. (2010)	UK	4	N	Y	Y	Y	Y
Lejeune et al. (2004)	France	2	Y	N	NA	NA	N
Lejeune et al. (2010)	France	2	Y	Y	Y	Y	N
Lejeune et al. (2014)	France	15	Y	Y	Y	Y	Y
Lew et al. (2017)	Australia	1	Y	Y	Y	Y	Y
Lew et al. (2018a)	Australia	9	Y	Y	Y	Y	Y
Lew et al. (2018b)	Australia	13	Y	Y	Y	Y	Y
Macafee et al. (2008)	UK	1	N	Y	Y	Y	N
McLeod et al. (2017)	New Zealand	1	Y	Y	Y	Y	N
Melnitchouk et al. (2018)	Ukraine	3	Y	Y	Y	Y	Y
Murphy et al. (2017)	UK	6	Y	Y	Y	Y	Y
Naber et al. (2019)	US	7	Y	Y	Y	Y	Y
O'Leary et al. (2004)	Australia	4	Y	Y	Y	Y	N
Parekh et al. (2008)	US	6	Y	Y	Y	Y	N
Peterse et al. (2021)	US	9	Y	Y	Y	Y	Y
Phisalprapa et al. (2019)	Thailand	2	Y	Y	Y	Y	Y
Pickhardt et al. (2007)	US	4	Y	Y	Y	Y	N
Pignone et al. (2011)	Australia	1	N	Y	Y	Y	N
Regge et al. (2009)	US	6	Y	Y	Y	Y	N
Salkeld et al. (1996)	Australia	1	N	Y	Y	Y	N
Sekiguchi et al. (2016)	Japan	3	Y	Y	Y	Y	N
Sekiguchi et al. (2020)	Japan	3	Y	Y	Y	Y	N
Senore et al. (2019)	Italy	3	Y	Y	Y	Y	Y
Sharaf and Ladabaum (2013)	US	6	Y	Y	Y	Y	N

(Continued)

Authors (Year)	Country/ Region	No. strategies	Report threshold	Report numeric estimates	Report sufficient figures	Report all results	Cost-effectiveness plane
Sharp et al. (2012)	Ireland	8	N	Y	Y	Y	Y
Shimbo et al. (1994)	Japan	7	Y	Y	Y	Y	Y
Song et al. (2004)	US	5	N	Y	Y	Y	Y
Sonnenberg and Delco (2002)	US	2	N	Y	Y	Y	N
Sonnenberg et al. (1999)	US	2	N	Y	Y	Y	N
Sonnenberg et al. (2000)	US	3	N	Y	Y	Y	N
Subramanian et al. (2009)	US	1	Y	Y	Y	Y	N
Tafazzoli et al. (2009)	US	11	N	Y	?	N	Y
Tappenden et al. (2007)	UK	5	N	Y	Y	Y	Y
Telford et al. (2010)	Canada	10	Y	Y	Y	Y	Y
Tsoi et al. (2008)	Asia	3	N	Y	Y	Y	Y
Van Der Meulen et al. (2018)	Netherlands	480	Y	Y	Y	N	Y
van Hees et al. (2014)	Netherlands	5	Y	Y	Y	Y	N
van Rossum et al. (2011)	Netherlands	2	N	Y	Y	Y	N
Vanness et al. (2011)	US	5	Y	Y	Y	Y	Y
Vijan et al. (2001)	US	7	N	Y	Y	Y	N
Vijan et al. (2007)	US	8	Y	Y	Y	Y	N
Wagner et al. (1991)	US	4	N	Y	Y	Y	N
Wang et al. (2012)	China	2	N	Y	Y	Y	Y
Wilschut, Hol, et al. (2011)	Netherlands	240	N	Y	N	N	Y
Wilschut, Habbema, et al. (2011)	Netherlands	288	Y	Y	?	N	Y
Wong et al. (2015)	Hong Kong	5	Y	Y	Y	Y	Y
Wu et al. (2006)	Taiwan	5	Y	Y	Y	Y	Y

NA, not applicable; ?, insufficient information.
Lew et al. (2018a). Benefits, Harms, and Cost-Effectiveness of Potential Age Extensions to the National Bowel Cancer Screening Program in Australia. Cancer Epidemiol Biomarkers Prev, 27(12), 1450-1461. Lew et al. (2018b). Evaluation of the benefits, harms and cost-effectiveness of potential alternatives to iFOBT testing for colorectal cancer screening in Australia. Int J Cancer, 143(2), 269-282.
The details of the assessment items can be found in the Methods section of this study.

Table 2-8 The detailed results according to Parts 2 & 3 of our quality assessment: the ICER calculation and the adequacy of strategies inclusion

Authors (Year)	Incremental comparisons[l]	Correct comparators	Both types of modalities included	Consistent schedule	All strategies in base-case analysis	No-screening included	Sufficient ICER range
Areia et al. (2019)	N	N	Y	Y	Y	Y	Y
Aronsson et al. (2017)	N	N	Y	Y	Y	Y	N
Azad et al. (2020)	N	Y	Y	N	Y	Y	N
Barouni et al. (2012)	N	Y	Y	Y	Y	Y	N
Barzi et al. (2017)	Y	Y	Y	N	Y	Y	N
Berchi et al. (2004)	Y	Y	Stool-based	Y	Y	N	N
Coldman et al. (2017)	N	N	Stool-based	Y	NA	Y	NA
Coretti et al. (2020)	Y	Y	Stool-based	Y	Y	Y	N
Dan et al. (2012)	Y	?	Y	N	Y	Y	Y
Dinh et al. (2013)	N	N	Y	N	Y	Y	N
Fitch et al. (2015)	N	Y	Image-based	Y	Y	Y	N
Flanagan et al. (2003)	?	?	Stool-based	Y	Y	Y	NA
Frazier et al. (2000)	Y	Y	Y	N	Y	Y	Y
Goede et al. (2013)	Y	Y	Stool-based	Y	Y	Y	Y
Goede et al. (2017)	Y	N	Stool-based	Y	Y	Y	Y
Greuter et al. (2016)	Y	N	Y	N	Y	Y	N
Greuter et al. (2017)	Y	Y	Stool-based	Y	Y	Y	N
Gyrd-Hansen (1998)	?	?	Stool-based	Y	NA	N	NA
Gyrd-Hansen et al. (1998)	Y	Y	Stool-based	Y	Y	Y	N
Hassan et al. (2008)	N	Y	Image-based	Y	Y	Y	N
Hassan et al. (2011)	N	Y	Y	N	Y	Y	N
Heitman et al. (2010)	N	N	Y	N	Y	Y	N
Heresbach et al. (2010)	N	N	Y	Y	Y	Y	NA
Huang et al. (2014)	N	Y	Stool-based	Y	Y	Y	N
Jahn et al. (2019)	Y	Y	Y	Y	Y	Y	N
Khandker et al. (2000)	N	Y	Y	N	Y	Y	Y
Knudsen et al. (2010)	Y	Y	Y	N	Y	Y	Y
Ladabaum and Mannalithara (2016)	N	N	Y	N	Y	Y	N

(Continued)

Authors (Year)	Incremental comparisons[1]	Correct comparators	Both types of modalities included	Consistent schedule	All strategies in base-case analysis	No-screening included	Sufficient ICER range
Ladabaum et al. (2004)	N	N	Image-based	Y	Y	Y	N
Ladabaum et al. (2013)	N	N	Y	N	Y	Y	N
Ladabaum et al. (2014)	N	N	Y	N	Y	Y	N
Ladabaum et al. (2019)	Y	N	Y	N	Y	Y	Y
Lansdorp-Vogelaar et al. (2010)	Y	Y	Y	N	N	Y	Y
Lee et al. (2010)	Y	Y	Y	Y	Y	N	N
Lejeune et al. (2004)	?	?	Stool-based	Y	N	Y	NA
Lejeune et al. (2010)	N	N	Stool-based	Y	Y	Y	N
Lejeune et al. (2014)	Y	N	Stool-based	Y	Y	N	Y
Lew et al. (2017)	Y	Y	Stool-based	Y	Y	Y	Y
Lew et al. (2018a)	Y	Y	Stool-based	Y	Y	Y	Y
Lew et al. (2018b)	Y	Y	Y	N	Y	Y	Y
Macafee et al. (2008)	Y	Y	Stool-based	Y	Y	Y	N
McLeod et al. (2017)	Y	Y	Stool-based	Y	Y	Y	N
Melnitchouk et al. (2018)	Y	Y	Y	N	Y	Y	N
Murphy et al. (2017)	Y	Y	Stool-based	Y	Y	N	N
Naber et al. (2019)	Y	Y	Y	N	Y	Y	N
O'Leary et al. (2004)	N	Y	Y	Y	Y	Y	N
Parekh et al. (2008)	N	N	Y	N	Y	Y	N
Peterse et al. (2021)	Y	N	Y	N	Y	Y	N
Phisalprapa et al. (2019)	Y	Y	Y	Y	N	Y	N
Pickhardt et al. (2007)	N	N	Image-based	Y	Y	Y	N
Pignone et al. (2011)	Y	Y	Stool-based	Y	NA	Y	N
Regge et al. (2009)	N	Y	Image-based	Y	Y	Y	Y
Salkeld et al. (1996)	Y	Y	Stool-based	Y	Y	Y	N
Sekiguchi et al. (2016)	N	N	Y	Y	N	Y	N
Sekiguchi et al. (2020)	N	N	Y	Y	Y	Y	N
Senore et al. (2019)	N	Y	Y	Y	N	Y	N
Sharaf and Ladabaum (2013)	N	N	Y	N	Y	Y	N

(Continued)

Authors (Year)	Incremental comparisons[ł]	Correct comparators	Both types of modalities included	Consistent schedule	All strategies in base-case analysis	No-screening included	Sufficient ICER range
Sharp et al. (2012)	N	Y	Y	Y	Y	Y	N
Shimbo et al. (1994)	N	Y	Y	N	Y	Y	Y
Song et al. (2004)	N	N	Y	N	Y	Y	Y
Sonnenberg and Delco (2002)	N	Y	Image-based	Y	Y	Y	N
Sonnenberg et al. (1999)	N	Y	Image-based	Y	Y	Y	N
Sonnenberg et al. (2000)	N	N	Y	N	Y	Y	N
Subramanian et al. (2009)	Y	Y	Y	Y	Y	Y	N
Tafazzoli et al. (2009)	Y	Y	Y	N	Y	Y	NA
Tappenden et al. (2007)	N	Y	Y	N	Y	Y	N
Telford et al. (2010)	Y	Y	Y	N	Y	Y	N
Tsoi et al. (2008)	N	N	Y	N	Y	Y	N
Van Der Meulen et al. (2018)	Y	Y	Image-based	Y	Y	Y	N
van Hees et al. (2014)	Y	Y	Image-based	N	Y	Y	N
van Rossum et al. (2011)	Y	Y	Stool-based	Y	Y	Y	N
Vanness et al. (2011)	Y	Y	Y	N	Y	Y	N
Vijan et al. (2001)	Y	Y	Y	N	Y	Y	Y
Vijan et al. (2007)	N	N	Y	N	Y	Y	Y
Wagner et al. (1991)	N	Y	Y	Y	Y	Y	N
Wang et al. (2012)	Y	Y	Image-based	Y	Y	Y	N
Wilschut, Hol, et al. (2011)	Y	Y	Stool-based	Y	Y	Y	N
Wilschut, Habbema, et al. (2011)	Y	Y	Stool-based	Y	Y	Y	NA
Wong et al. (2015)	N	N	Y	Y	Y	Y	N
Wu et al. (2006)	N	N	Y	N	Y	Y	N

ICER, incremental cost-effectiveness ratio; NA, not applicable; ?, insufficient information.

ł Non-incremental comparisons include cost-effectiveness ratio without comparing to any intervention and ratio compared to no-screening.

Lew et al. (2018a). Benefits, Harms, and Cost-Effectiveness of Potential Age Extensions to the National Bowel Cancer Screening Program in Australia. Cancer Epidemiol Biomarkers Prev, 27(12), 1450-1461. Lew et al. (2018b). Evaluation of the benefits, harms and cost-effectiveness of potential alternatives to iFOBT testing for colorectal cancer screening in Australia. Int J Cancer, 143(2), 269-282.

The details of the assessment items can be found in the Methods section of this study.

Association of quality performance with study characteristics

All the results regarding the correlation of examined quality items and study characteristics are tabulated, including their year of publication (Table 2-9), the use of cost-effectiveness planes (Table 2-10), and their country (Table 2-11).

The year of publication was not related to most of the items assessed in this review. Only the inclusion of a cost-effectiveness plane was found more often in more recent CEAs. Studies that featured cost-effectiveness planes tend to have both types of screening modalities and consistent screening schedules. These studies also tend not to confuse ratios compared to no-screening and ICERs. Notably, this review found significant regional differences in these items as well. For instance, Asian (n = 9, 73%) and North American (n = 20, 61%) CEAs tend to report ratios based on the comparison to no-screening as ICERs, while Australasian (n = 1, 14%) and European CEAs (n = 9, 35%) do not. North America has the highest proportion, 71% (n = 24), of CEAs including both types of screening modalities, while it was the only region that less than half studies (n = 12, 35%) included consistent screening schedules. Conversely, only 29% (n = 2) of Australasian CEAs include both types of modalities, although it has the highest percentage (n = 6, 86%) of CEAs analysing consistent schedules. Furthermore, all the CEAs excluding a no-screening option were European (*p-value* < 0.05).

Table 2-9 The association between the quality and the year of publication

	The year of publication											
	All	**<=2000**		**2001-2005**		**2006-2010**		**2011-2015**		**2016-2020**		**P-Value**
		N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	
Completeness of reporting												
Report both cost and effect estimates												0.101
Yes	75	8	89	6	75	17	94	20	100	24	96	
No	5	1	11	2	25	1	6	0	0	1	4	
Sufficient significant figures												0.125
Yes	70	8	100	6	100	16	100	16	84	24	100	
No	3	0	0	0	0	0	0	3	16	0	0	
All the strategies												0.636
Yes	67	7	88	6	100	16	94	16	80	22	92	
No	8	1	13	0	0	1	6	4	20	2	8	
Cost-effectiveness plane												0.023
Yes	47	4	44	1	13	10	56	13	65	19	76	
No	33	5	56	7	88	8	44	7	35	6	24	
ICER calculation												
Incremental comparisons[ł]												0.259
Yes	39	3	38	2	33	7	39	10	50	17	68	
No	38	5	63	4	67	11	61	10	50	8	32	
Correct comparators												0.622
Yes	49	7	88	4	67	10	56	13	68	15	60	
No	27	1	13	2	33	8	44	6	32	10	40	
Adequacy of strategies inclusion												
Both stool and image-based modalities included												0.443
Yes	47	5	56	3	38	13	72	10	50	16	64	
No	33	4	44	5	63	5	28	10	50	9	36	
Consistent screening schedule												0.629
Yes	47	5	56	6	75	8	44	12	60	16	64	
No	33	4	44	2	25	10	56	8	40	9	36	
No-screening option included												0.824
Yes	75	8	89	7	88	17	94	19	95	24	96	
No	5	1	11	1	13	1	6	1	5	1	4	
No additional strategies in sensitivity analysis												0.42
Yes	72	8	100	7	88	17	94	19	100	21	88	
No	5	0	0	1	13	1	6	0	0	3	13	
The range of efficient frontier encompasses the threshold												0.738
Yes	18	3	38	2	33	4	25	3	16	6	25	
No	55	5	63	4	67	12	75	16	84	18	75	

ICERs, incremental cost-effectiveness ratios.

ł Non-incremental comparisons include cost-effectiveness ratio without comparing to any intervention and ratio compared to no-screening.

The details of the assessment items can be found in the Methods section of this study.

Table 2-10 The association between the quality and the use of the cost-effectiveness plane

	Cost-effectiveness plane					P-Value
	All	Y		N		
		N	*%*	*N*	*%*	
Completeness of reporting						
Report both cost and effect estimates						0.644
Yes	75	30	91	45	96	
No	5	3	9	2	4	
Sufficient significant figures						1
Yes	70	29	97	41	95	
No	3	1	3	2	5	
All the strategies						0.134
Yes	67	29	97	38	84	
No	8	1	3	7	16	
ICER calculation						
Incremental comparisons[†]						0.008
Yes	39	10	32	29	63	
No	38	21	68	17	37	
Correct comparators						0.333
Yes	49	18	58	31	69	
No	27	13	42	14	31	
Adequacy of strategies inclusion						
Both stool and image-based modalities included						0.043
Yes	47	15	45	32	68	
No	33	18	55	15	32	
Consistent screening schedule						0.033
Yes	47	24	73	23	49	
No	33	9	27	24	51	
No-screening option included						0.399
Yes	75	32	97	43	91	
No	5	1	3	4	9	
No additional strategies in sensitivity analysis						1
Yes	72	29	94	43	93	
No	5	2	6	3	7	
The range of efficient frontier encompasses the threshold						0.186
Yes	18	5	17	13	30	
No	55	25	83	30	70	

ICERs, incremental cost-effectiveness ratios.

† Non-incremental comparisons include cost-effectiveness ratio without comparing to any intervention and ratio compared to no-screening.

The details of the assessment items can be found in the Methods section of this study.

Table 2-11 The association between the quality and the researched region

	All	Area								P-Value
		Asia		Australasia		Europe		North America		
		N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	
Completeness of reporting										
Report both cost and effect estimates										0.732
Yes	75	11	100	7	100	25	89	32	94	
No	5	0	0	0	0	3	11	2	6	
Sufficient significant figures										0.683
Yes	70	10	91	7	100	23	96	30	97	
No	3	1	9	0	0	1	4	1	3	
All the strategies										0.379
Yes	67	10	91	7	100	20	80	30	94	
No	8	1	9	0	0	5	20	2	6	
Cost-effectiveness plane										0.123
Yes	47	7	64	3	43	21	75	16	47	
No	33	4	36	4	57	7	25	18	53	
ICER calculation										
Incremental comparisons[ł]										0.021
Yes	39	3	27	6	86	17	65	13	39	
No	38	8	73	1	14	9	35	20	61	
Correct comparators										0.061
Yes	49	5	50	7	100	19	73	18	55	
No	27	5	50	0	0	7	27	15	45	
Adequacy of strategies inclusion										
Both stool and image-based modalities included										0.02
Yes	47	9	82	2	29	12	43	24	71	
No	33	2	18	5	71	16	57	10	29	
Consistent screening schedule										0.002
Yes	47	7	64	6	86	22	79	12	35	
No	33	4	36	1	14	6	21	22	65	
No-screening option included										0.028
Yes	75	11	100	7	100	23	82	34	100	
No	5	0	0	0	0	5	18	0	0	
No additional strategies in sensitivity analysis										0.303
Yes	72	9	82	6	100	25	93	32	97	
No	5	2	18	0	0	2	7	1	3	
The range of efficient frontier encompasses the threshold										0.214
Yes	18	2	18	3	43	3	13	10	32	
No	55	9	82	4	57	21	88	21	68	

ICERs, incremental cost-effectiveness ratios.

ł Non-incremental comparisons include cost-effectiveness ratio without comparing to any intervention and ratio compared to no-screening.

The details of the assessment items can be found in the Methods section of this study.

Discussion

Brief findings

This review identified two clear methodological issues in CEAs of CRC screening published to date. First, the misinterpretation of ICERs is widespread as this review finds less than half of the CEAs reviewed reported ICERs that are in accordance with the ratio's standard definition. Second, the omission of potentially relevant CRC screening strategies is also prevalent as this review finds all CEAs except one (Areia et al., 2019) had at least one issue in the inclusion of screening strategies.

The problem of reporting incorrect ICERs primarily arises for two reasons: those of inappropriate comparisons between strategies and those of inappropriate presentation. Examples of the first type include comparison of strategies to dominated strategies including no screening. Examples of the second type are the presentation of CERs that do not correspond to the definition of the ICER as ICERs. It is possible many analysts may appreciate that the ICER is the required ratio to inform decision-making but still also use the term ICER to denote other ratios for some reason.

Incremental comparisons

A large proportion of the reviewed CEAs reported ratios estimated on the basis of comparisons to no screening as ICERs, despite it long being clearly recognised that the ICER is the appropriate metric (Cohen & Reynolds, 2008; Hoch & Dewa, 2008). Ratios on the basis of comparisons to no screening have been termed ACERs and their use can mislead a decision-maker into accepting a cost-ineffective strategy (Torrance et al., 1996). This is especially likely if the decision-maker is not familiar with the distinction between ACERs and ICERs. Similarly, this review found many studies reported both correctly calculated ICERs and incorrectly calculated ICERs based on alternative comparisons. The review indicates that it would be beneficial if analysts wish to report ratios other than ICERs that they take care not to label such alternative ratios as ICERs to avoid confusion.

Results reporting

The completeness of reporting has implications for transparency. Not all the reviewed CEAs reported costs and effects for all simulated strategies. This withholds potentially relevant information from readers. Some CEAs presumably do not report complete results due to the inclusion of a large number of strategies. This thesis suggests future studies report all their results numerically in appendices.

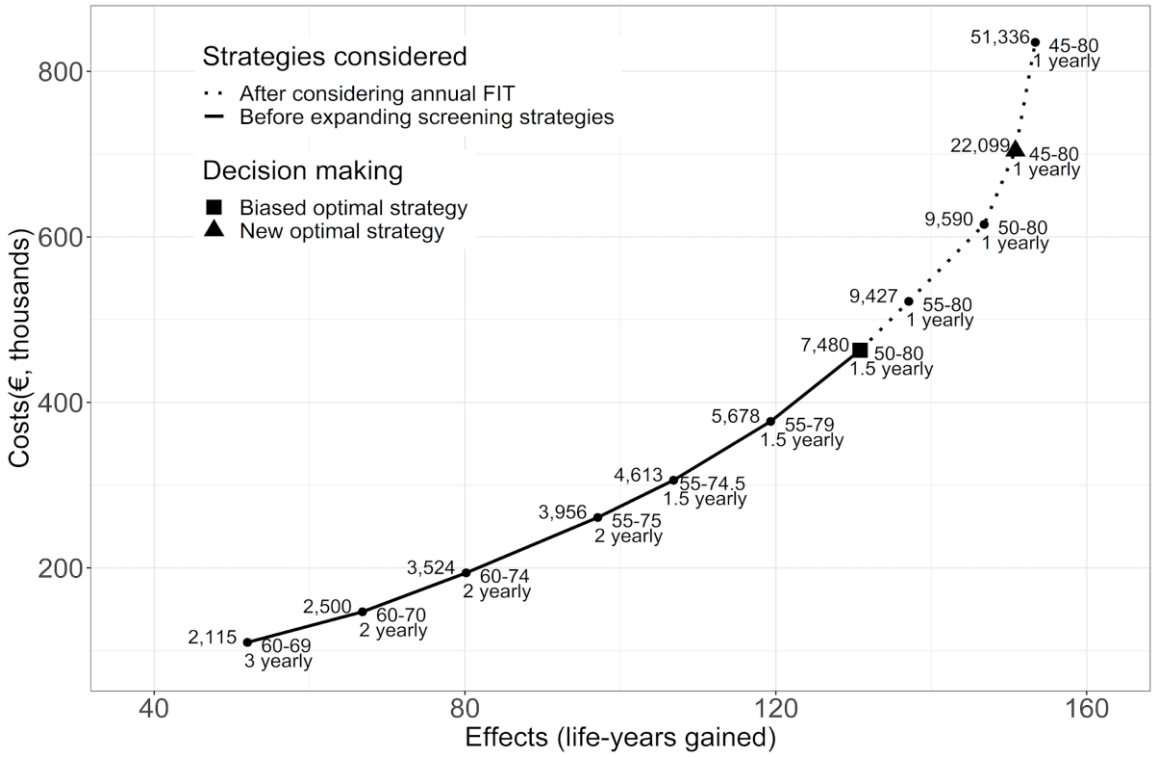
Misinterpretation of CERs in the CRC literature appears to stem, in part at least, from the use of cross-tabulation tables in which all strategies are compared against each other, such as that reported by Sonnenberg et al. reproduced in Figure 2-3. Such tables appear particular to the CRC literature as they are uncommon in CEAs of screening for other cancers. This thesis also suggests that analysts stop reporting cross-tabulation tables as it seems these are more misleading than helpful. Instead, analysts should follow existing clear guidance on how to tabulate cost-effectiveness results offered by either Siegel et al. (1996) or Prosser et al. (2016).

Strategy choice

The scope of strategy choice significantly impacts how well researchers address their research questions or objectives. One possible reason for the exclusion of strategies in CEAs is the specific aim of the study. CEAs with narrower objectives may focus solely on comparing screening modalities while intentionally omitting variations in screening schedules. In contrast, this study provides a broader perspective on the issue of screening strategy choice in CEAs, emphasising the pursuit of the most optimal strategy from the array of options. Appendix 1 of this thesis systematically reviewed the CRC screening strategies analysed in European CEAs and examined whether the strategy choices aligned with the objectives of these CEAs. Appendix 1 found that approximately 72% (n = 28) of CEAs had broader objectives, aiming to identify the optimal strategy for policy purposes. Among these studies with broader objectives, more than half (n = 16, 57%) simulated five strategies or fewer. This limited scope of strategies is incongruent with the objectives of these studies. The findings suggest that CEAs with restricted strategy choices may not fully address their research questions.

It seems likely that many CEAs choose strategies from those examined previously or defined within clinical guidelines without adding other strategies as potentially relevant policy options (Lansdorp-Vogelaar et al., 2011; Patel & Kilgore, 2015). While an understandable approach, without the inclusion of any additional strategies such analyses therefore have limited the analysis at the outset. For example, this approach appears to have limited many analyses within the CRC review from considering potentially relevant schedules such as those starting below age 50 and ending after age 70. While varying screening age range has been investigated by previous analyses, there appears scope for further examination of the benefits of expanding screening schedules (Imperiale et al., 2018; Mannucci et al., 2019).

Figure 2-4 The cost-effectiveness frontier from the study of Goede et al. (2013)



FIT, immunochemical faecal occult blood test. The labelled numbers above and below the efficient frontier are the incremental cost-effectiveness ratio (ICER) and the screening schedule (age range and interval) for each strategy respectively. ICERs are expressed in euros (€) per life-year gained. Optimal strategy is decided by the threshold of €35,916 per life-year gained which is the median of cost-effectiveness analyses (CEAs) studied the same country in our review because Goede et al. (2013) did not report the threshold estimate. This figure illustrates the impact of excluding annual FIT based on the example from Goede et al. (2013). If the analysis had omitted annual FIT, it would have resulted in estimating a range of ICER that do not fall within the threshold value. Consequently, the analysis might have mistakenly concluded that a less effective and lower-intensity strategy with a lower ICER was the optimal choice.

Many European CRC screening guidelines recommend biennial FIT (Schreuders et al., 2015), and many of the studies reviewed in this thesis did not consider a shorter annual interval. All of the studies simulating a large range of screening strategies found annual

screening to be cost-effective (Goede et al., 2017; Goede et al., 2013; Gyrd-Hansen, 1998; Wilschut, Habbema, et al., 2011; Wilschut, Hol, et al., 2011), but this review only found 63% (n = 31) of CEAs analysing FIT (n = 49) simulated annual screening. The study presented in Appendix 1 of this thesis also yielded similar findings. Among the 39 studies, 37 primarily examined biennial screening intervals, and only 13 studies considered annual screening. Appendix 1 further compared the strategies considered in CEAs to the current screening policies in European countries. Remarkably, despite all 13 of these studies recommending annual screening for optimal cost-effectiveness, the prevailing screening policy in stool-based programmes involves biennial screening in 25 out of 26 European countries.

Figure 2-4 demonstrates the consequence of omitting annual FIT using the example of Goede et al. (2013). Had that analysis excluded annual FIT, it would have estimated a range of ICERs not encompassing the threshold and would have mistakenly concluded a less effective, lower-intensity strategy with lower ICER was optimal. The example of Goede et al. (2013) would choose the strategy of aged 50-80 every 1.5 years, instead of a more frequent strategy with annual screening between aged 45 and 80. Similarly, among the 57 CEAs that included image-based modalities, only half (n = 29, 51%) considered five-yearly screening intervals, meaning those that did not consider such short intervals might have overlooked more effective strategies (Van Der Meulen et al., 2018; van Hees et al., 2014).

There may be a difference of opinion between opinions from the perspectives of clinical experts and health economists. Clinical guidelines tend to suggest a narrower selection of strategies as comparators in the CEA, compared to the recommendations of health economists. A possible reason is that many health economists are aware of the need to simulate both the strategies of interest and the comparators with lower intensities as detailed in Torrance et al., (1996). This need might not be recognised by clinicians. In addition, clinicians might be uncomfortable departing from those intervals directly considered in trials, whereas health economists might be more willing to consider extrapolations based on assumptions of disease progression in order to consider what was not trialled. Health economists, compared to clinicians, more recognise the advantages of the model that can simulate more strategies than those considered in trials.

Accordingly, this thesis suggests that analyses should not be constrained by current clinical guidelines when specifying what strategies to compare.

This thesis acknowledges infrequent strategies, e.g., with long screening intervals, might be chosen due to capacity constraints. For example, a fixed number of colonoscopies is available in a realistic setting. Nevertheless, it is still worth simulating an unconstrained scenario despite resource constraints. This thesis suggests CEAs illustrate such constraints in their simulation and should not excessively limit their range of strategies at its outset. For instance, an example of good practice is Wilschut et al (2011) which simulated screening with gFOBT or FIT every 1, 1.5, 2, and 3 years and discussed the ideal screening strategies for different colonoscopy capacity levels. It might also be worth studying different combinations of modalities or triage tests to reduce the colonoscopy burden. For example, the triage test can be another biomarker in the same sample as taken for FIT (Greuter et al., 2020).

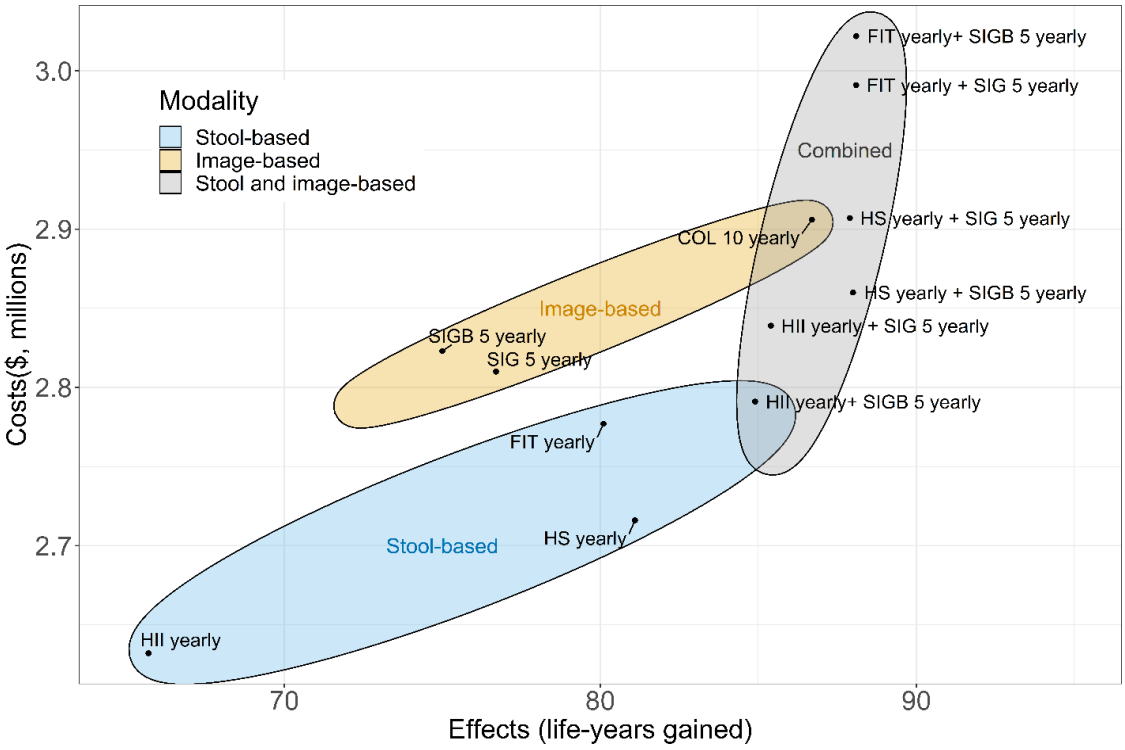
Possible strategy omission

Many studies in this review reported ICERs that are exclusively above or below the cost-effectiveness threshold. Only a small number of studies feature large ranges of strategies that have relatively complete frontiers with ICERs which extend from below threshold and then rise above. A potential oversight in identifying the optimal strategy may arise when the range of ICERs does not encompass the threshold value. This is because the conventional rule of decision-making in healthcare is to select the strategy with the highest ICER that falls just below the threshold as the optimal choice. Given the inherent uncertainty surrounding the threshold estimate, it is advisable for CEAs to ensure that their ICER range encompasses the entire spectrum of potential threshold values. If all the ICERs are above or below the threshold, CEAs can enhance their analysis by either constraining or expanding the scope of screening intensity explored in the studies. The example of strategy expansion based on the observation of prior results is further illustrated in Study 3 of this thesis. This thesis notes, however, that a large range of ICERs on the frontier is not a guarantee of a complete set of strategies, but that it offers at least some reassurance. The ideal analysis includes a broad screening age range and intervals, and the simulation of the range of different screening modalities available.

This review only found one CEA featuring all seven CRC screening modalities considered (Telford et al., 2010). This study, however, did not include a broad range of alternative age ranges or screening intervals. Therefore, although it was rich in terms of modality variation, it did not feature comprehensive intensity variation. Conversely, the studies that featured the most comprehensive intensity variation only included stool-based tests. The omission of image-based tests and strategies that combine both modalities means they may not have fully identified all relevant alternatives despite simulating hundreds of strategies.

Studies examining both image and stool-based tests generally found stool-based modalities were less costly but less effective compared to image-based modalities. Though some caution is required with this generalisation owing to the lack of variation in intensity in studies considering both types of modalities. Furthermore, the finding that image-based modalities are optimal also has to be considered in the context of the findings of Kundsen et al. (2010) which are reproduced in Figure 2-5. It is notable that includes strategies that combined both stool and image-based testing, such as annual FIT testing accompanied by sigmoidoscopy every five years, can be the preferred strategies as these dominate stool only or image only screening approaches (Knudsen et al., 2010). The lack of investigation of such combined strategies might be a problem for the CRC literature as these are very little investigated yet the available evidence indicates they are superior.

Figure 2-5 Screening modalities simulated by Knudsen et al. (2010) (partial results)



COL, colonoscopy; FIT, immunochemical faecal occult blood test; HII, Hemoccult II; HS, Hemoccult SENSA; SIG, sigmoidoscopy without biopsy; SIGB, sigmoidoscopy with biopsy. In this example, the strategies with stool-based modalities have lower cost estimates. The combination of both types of screening modalities tends to have higher costs and effectiveness.

Heterogeneity of reporting

This review found pronounced heterogeneity in the presentation of the CEAs reviewed. Reporting heterogeneity exists in the description of the model type, the details of the simulated strategies, the reporting format for costs and health effects, and the terminology used for CERs. Despite clear guidance, CEAs of screening tend to have incomplete and inconsistent reporting, resulting in a barrier to accessing the key information. Readers require more time to understand the study design and the results of CEAs due to inconsistent reporting format. A more fundamental problem of transparency occurs when CEAs fail to disclose important information about their methods and results.

Incomplete reporting of the model type is observed in the literature on CEAs of screening. When CEAs of screening describe their simulation model, they usually fail to state important attributes of model structure. For example, microsimulation is a general term for individual simulation, in contrast to cohort simulation. Describing a model as a microsimulation does not describe the model structure, properties of the health states or

whether time is handled discretely or continuously. Without more detailed reporting or open sourcing of the model it is hard for readers to comprehend the study design and its corresponding assumptions. It would be helpful if modellers could give more details in their simulation model.

An important observation from the review is that the description of strategies is not standardised. A large range of strategies can be considered with possible variation in multiple aspects, such as screening modalities and screening intensities. Although some guidelines have suggested CEAs disclose the details of the strategies considered (more details can be found in the introduction chapter of this thesis), without guidance in reporting format, it still typically still requires reader effort to clearly identify the details of these strategies. Notably, it was observed that many CEAs considering a large number of strategies employ a similar format of reporting that is clear regarding the strategy characteristics. Such studies typically defined strategy names that indicated the screening age range, interval and primary screening modality. For example, Wilchut et al (2011) use the strategy abbreviation of *FIT 50 GS, 45-80, 1 year* to denote FIT screening with a cut-off of 50 ng/ml with guidelines surveillance for positive patients for individuals aged 45 to 80 with annual screening. Furthermore, such studies might systematically apply the same choice of screening intervals and age ranges to an alternative modality, such as FOBT. In doing so these analyses hold the choice of intervals and age ranges constant between the analyses clarifying what differences are due to the differences in modalities. Naturally the same does not apply to comparisons of very different modalities such as FIT and colonoscopy as primary screening tests.

There are further variations in the way costs and effects are reported. Studies variously report costs and effects estimates of strategies vs no-screening, vs one specific screening strategy or simply report the total estimated lifetime costs and effects without referencing back to a baseline no-screening scenario. Accordingly, there is no consensus on the standard reporting format for the main cost-effectiveness outcomes. This lack of standardization can present a challenge when interpreting results as the format varies between studies. A particular challenge is presented by those studies reporting total lifetime costs and effects without reference to a no screening strategy. This is problematic as it does not permit a baseline demonstration that any strategy is at a minimum cost-effective relative to no screening. Sonnenberg et al. (1999), for example, provided the

total costs per life-year saved for each screening strategy, which were determined based on the crude costs of each intervention without comparison to any other strategy or reference to the no-screening scenario.

Problems in reporting of costs and effects estimates also relate to the various formats results are reported in including diagrams, tables and within manuscript texts. Some studies primarily report estimates graphically in cost-effectiveness planes, while others may use tables but not report results for all estimates and others may only report ICERs without reporting costs and effects. Such non-standardised reporting makes it time consuming to extract results when reviewing CEAs. Notably, CEAs only employing diagrams to present cost-effectiveness estimates fail to provide detailed numerical outcomes, which are necessary for the verification of ICER calculations. Furthermore, such graphical reporting can limit the identification of specific strategies, especially those behind the efficient frontier as labelling all strategies is typically infeasible when the number of strategies considered is large. Even when results are reported numerically, they need to be reported to sufficient significant figures to enable replication of the reported ICERs. Analysts may overlook what is an appropriate number of significant figures to report if they do not consider the incremental differences between strategies which determine ICER estimates.

The inconsistent use of terminology regarding the ICER is apparent in the screening CEA literature. The thesis interprets the ICER in accordance with its conventional use as the cost-effectiveness metric used for identifying the optimal intervention within a specific cost-effectiveness threshold. Accordingly, it refers to the ICERs calculated based on the comparisons among the strategies on the efficient frontier which rules out the simple and extended dominated. This thesis found some CEAs of CRC screening reported ICERs based on the comparison to the no-screening strategy which refers to the concept of the ACER. This misleading usage of “ICER” means readers cannot properly address the meaning of cost-effectiveness results, and presents difficulties when attempting to compare results across CEAs. This thesis suggests confusion can be avoided if modellers use appropriate terminology for presenting different cost-effectiveness results. In addition to correctly labelling ICERs and ACERs, other ratios should be labelled with CER, preferably noting its corresponding comparator against which costs and effects have been compared against.

Existing checklists for economic evaluation

Existing reviews of CRC screening CEAs have used structured checklists to examine the inclusion of strategies and have found most CEAs to be of high quality (Hanly et al., 2012; Mendivil et al., 2019; Ran et al., 2019). The review presented within this thesis contradicts those conclusions as most CEAs do not explore sufficient screening strategies. The explanation of these contradictory findings is that although many checklists have questions regarding the choice of alternative strategies, they address the transparency of reporting, not the appropriateness of the choice of alternatives compared. Accordingly, a CEA can satisfy the relevant criteria in both checklists if it provides a rationale for the choice of alternatives considered, irrespective of whether that rationale is sound or not.

Assessing all possible screening strategies is infeasible due to data and computing capacity constraints. Nevertheless, a good CEA should still include a sufficiently large range of strategies to avoid the choice of comparators becoming a source of significant bias. This thesis therefore suggests analysts simulate cost-effectiveness in an iterative fashion, using an initial analysis to identify what strategies are most likely to be relevant and then expand the analysis to strategies with potential for being cost-effective. For example, if screening continues to be effective and cost-effective at an upper screening age of 65, then it is likely worthwhile incrementally increasing the simulated stop age until the ICER exceeds the threshold. Although such an iterative approach runs counter to guidelines of economic evaluation which recommend determining the research question at the outset, this thesis believes it is more suitable in a screening context where the relevance of alternative strategies cannot meaningfully be identified before simulation is undertaken. Details on how to expand the inclusion of strategies with the iteration method can be found in the third study of this thesis.

Limitations

A limitation of the review presented in this thesis is that it is predicated on the assumption that a CEA attempts to find the optimally cost-effective strategy among all feasible strategies. Individual CEAs may address more specific questions relevant to their particular context, such as what is the optimal intensity of a given screening modality or will shifting to a new testing modality be cost-effective. Such specific questions differ

from the overall goal implicit in most cost-effectiveness analyses of finding the overall most cost-effective intervention strategies. Accordingly, the overall critique of the reviewed studies arguably should not apply in all cases.

The observations made in this review are confined to the CEAs of CRC screening that are included in this analysis. The review exclusively focuses on real-world screening services, omitting discussions on hypothetical screening methods, clinical diagnostic testing, and post-treatment surveillance. Additionally, CEAs designed for subgroup analysis in screening are not reviewed. Certain CEAs likely aim to differentiate screening intensity among distinct risk groups, a consideration that is absent from our discussion. Furthermore, this review excludes CEAs that report different measurements of health economic outcomes, such as LY gained per colonoscopy, limiting our discussion on varying CEA objectives in screening. This review also disregards CEAs within screening promotion programs designed to enhance adherence, thus impeding our ability to discuss the relationship between adherence considerations and strategy integration in CEAs. Notably, a possible rationale for omitting consideration of short screening intervals could be attributed to exceedingly low adherence rates. Moreover, our search was confined to English-language studies, which leaves the generalizability of our findings to other literature uncertain. Naturally, the conclusions drawn here do not extend to studies that were not included in our analyses.

A limitation of this review is the absence of a validated assessment tool specifically designed for strategy inclusion in CEAs of screening. The development of the assessment tool in this study primarily draws from literature observations and the relevant field experience of Dr. James O'Mahony, who has published several similar reviews in other cancer contexts (Fabbro et al., 2022; O'Mahony, 2014, 2015, 2019). Unfortunately, there has been no collaborative effort within the international community to create a checklist that would ensure the applicability of the listed evaluation criteria across diverse settings in CEAs. In addition, it is difficult to objectively appraise if a study includes enough strategies. Although this review tried to devise objective criteria that are assessed quantitatively when creating our quality assessment tool, a degree of subjective judgment remains.

Finally, the assessment conducted in this study exclusively focuses on identifying cost-effective strategies, particularly regarding the methodology used for comparing strategies and calculating ICERs, as well as the selection of strategies for inclusion. This study does not delve into other quality aspects of economic evaluation, such as model structure, assumptions about the disease process, measurement and valuation of outcomes, and the impact of uncertainty.

Future research

This thesis found existing checklists and guidelines for economic evaluation are not completely adequate in the context of screening. This is understandable if such guidance has been drafted primarily in the context of therapeutic interventions rather than the specific context of screening. It is therefore not surprising that screening-specific issues that relate to the choice of intervention strategies such as the range of screening intervals and age ranges are somewhat inadequately addressed by those guidelines. Similarly, issues of risk stratification and intervention adherence (screening uptake rates) also apply to screening in specific ways and these are also not well-reflected in existing CEA guidelines. This is then reflected in the review's findings regarding the shortcomings in results reporting, strategy choice, and ICER estimation. There is a need for further research to examine the appropriate application of existing guidance in CEAs of screening. Additionally, it is important for the international community to develop specific guidelines and evaluation checklists tailored for screening. The collaboration of experts can extend the aspects of screening CEAs that were not discussed in this thesis, such as the consideration of adherence rates, and improve the generalisation of the guidelines.

3. STUDY 2 A PEDAGOGICAL MODEL FOR CEA OF SCREENING

This study develops a simple, open-source modelling framework to illustrate the effect of the choice of relevant strategies and model parameters on the simulation of efficient frontiers. The first study of this thesis identified a methodological problem of under-identifying cost-effective strategies in the CEAs of CRC. This is likely because there is insufficient understanding of the relationship between the model inputs and the cost-effectiveness of strategies. The present study uses the model to illustrate how the position of the efficient frontier changes with variations in key parameters, including the screening test performance, disease history, and cancer incidence. Additionally, this study employs comparative statics to present the efficient frontiers.

Methods

This section described a pedagogical model of screening. This simplified microsimulation model is coded in R (version 4.2.1) and comprises approximately 730 lines including markup. The complete model code and its specification are available for all to access freely on GitHub (https://github.com/yishu-lin/Pedagogical-CEA-Model-of-Screening.git). The additional step-by-step guide can be found in the appendix of this thesis.

Model overview

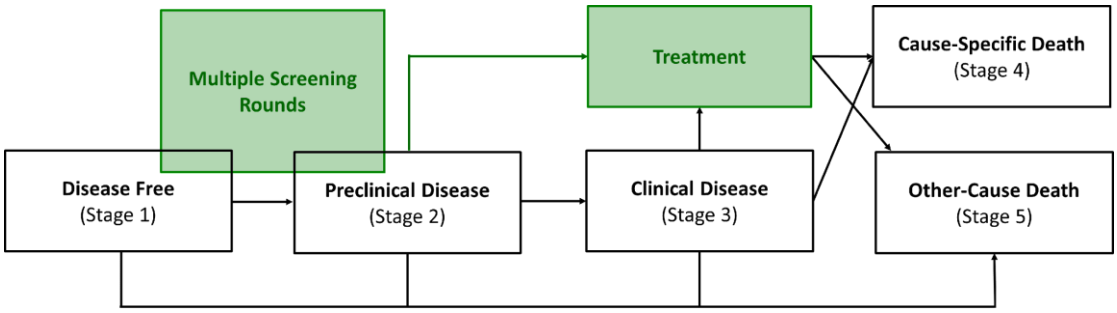
This study first provides a broad outline of the modelling approach before giving a detailed description in the following section on pseudo-code. This model was designed for illustrative purposes and does not represent any specific disease or intervention, though it is broadly conceived in the context of cancer in which there is a non-communicable preclinical disease that can be screened for at multiple points over an individual's lifetime. This model is deliberately more abstract than the applied models typically used in applied CEAs.

A single lesion model is adopted, meaning that individuals can only develop one instance of disease per lifetime. The disease can be treated, either upon clinical presentation or screen detection. Only one treatment is assumed available for the disease. Treatment is not assumed to be perfectly effective as not all patients survive, and those who survive

are assumed to have no long-term morbidity. This framework also assumed no disease recurrence.

This individual-level DES depicts a natural history of disease for a single lesion with five health states and two interventions (Figure 3-1). All individuals start in the perfect health state but are at risk of disease. In the absence of disease, individuals die of other causes, the timing of which is determined by assumed life tables. Individuals in the preclinical state have disease but have not been diagnosed and suffer no symptoms. Their disease is detectable by screening. Individuals can be diagnosed after either symptomatic presentation or screen detection, at which point they enter a clinical state and start treatment. This model assumes treatment has a higher probability of success if disease is detected in the preclinical state. If individuals are cured, there is no further treatment and survivors are excluded from future screening activity. In the case of treatment failure, individuals die at the same point in time as if no treatment occurred, i.e., there is no longevity benefit. In the case of treatment success, death occurs at the time of other cause death as determined by the assumed life tables. The model simulates all health outcomes for each individual until death.

Figure 3-1 Model diagram



The model that developed by this study includes five health states (transparent squares) and considers two possible interventions (green squares). The arrows are all single-directional, meaning individuals can only move to the next health state, as we do not consider disease recurrence, and deaths (Stage 4 and Stage 5) are absorbing states.

The model employs a vectorised approach, meaning that large vectors are used to record the age of entry and exit of specific states corresponding with each element in the vector corresponding to simulated individuals and the model operations are, wherever possible, applied over these vectors. The vectorised approach aids efficiency of the model and minimises the iterative use of for loops.

This study makes several assumptions to both ensure reproducible results across model iterations and reduce stochastic (first order) uncertainty. Each simulation starts with a random number seed. The use of common seeds permits holding the natural history of each individual constant across simulations. Other sample seeds are used to ensure a fair comparison among the screening strategies. For instance, this model applied the same random seeds for individuals for common age-specific screening moments in different simulations. This assumes the probability of detecting a true positive at a given screening moment will be the same across separate simulations featuring the same screening moment. Similarly, this model also assumed the same random numbers when simulating treatment success from screen-detected and symptomatically-presenting disease. This is to ensure any given individual has better outcomes from screen-detected disease and treatment outcomes are comparable across strategies. The same random seeds are used for the probability of cure for symptomatic presentation across iterations.

Pseudo-code: the model structure

Simulation of natural history

First, the model simulates the natural history of each individual. The ages at entering the preclinical state and other-cause death are independently drawn from the probability distributions of disease incidence and life tables. Both distributions are defined by the incidence probability and the survival rate for the user-defined age groups. The sojourn time of the preclinical and clinical stages are assigned from user-defined distributions. Users can choose from uniform, exponential, or Weibull distributions. An individual's age exiting a given health state is determined by their age at entry plus the sampled sojourn time in that state. An individual's all-cause death age is the minimum of the cause-specific and other-cause death ages.

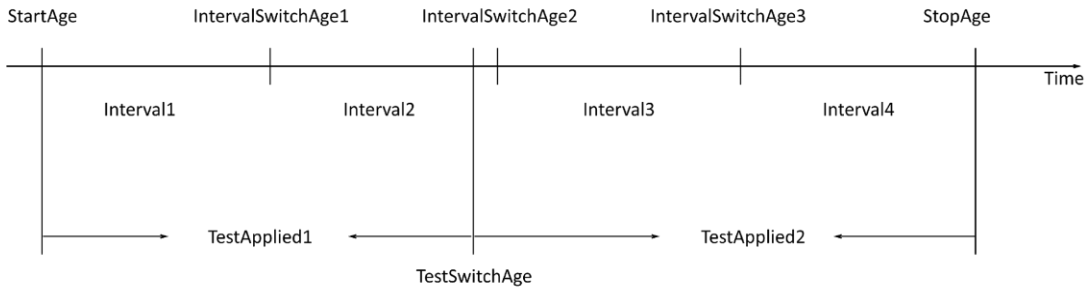
Simulation of screening strategies

The model includes a primary screening test, the sensitivity and specificity of which can be adjusted as can the interval between screens and the start and stop age. The model assumes a disutility from primary screening to account for the QoL loss due to the time and effort associated with undergoing a screening test. The model permits the simulation of alternative sets of test performance assumptions, corresponding to alternative primary screen modalities. Which modalities are applied and the screening interval can be varied

over the course of an individual's screening programme (Figure 3-2). The model does not explicitly simulate a triage test but does assume all the positives receive triage and only true positives access treatment. This model considers the harm the triage brings, including extra costs and disutility.

The screening schedules are created as approximations of the target age ranges and intervals. For example, triennial screening with a start age of 40 and stop age of 60 is approximated by 8 screens between the ages of 40 and 61. This approximation is achieved by holding the starting age and interval fixed but choosing the stop age that gives the closest approximation to the target stop age. Where two alternative stop ages can approximate a given target stop age this framework specifies the higher of the two. For example, while both 68 and 72 could approximate 70 our model applies 72.

Figure 3-2 Screen schedule



These are the parameter names used in the simulation to define screening strategies. StartAge is the age of starting the screening programme and StopAge is the end of the screening. While the default setting for the model is to use the same screening interval length and test modality throughout the whole screening programme, these can be varied at specific ages. IntervalSwitchAge is the age at which changes in the length of the screening intervals changes at, which can be changed up to three times. TestSwitchAge is the age that changes screening modality from TestApplied1 to TestApplied2.

Cost and effects estimates

This model considered four types of costs: primary screen; triage; early treatment; and, late treatment. The primary screen cost is calculated based on the total number of screens conducted regardless of their outcomes. The model applies the cost of triage testing to all those primary test positives (including false positives). Early treatment is received by true positive patients identified by screening. Conversely, late treatment is received by those individuals presenting symptomatically. The treatment costs occur when individuals undergo treatment, which is assumed to occur at a single point in time per patient.

The model includes QoL adjustment. Both screening and triage incur QoL losses. QoL losses are also applied to treatment for screen-detected and symptomatic disease on a once-off basis. This model assumed no QoL decrement for being in the preclinical state including after screen detection, though this model assumed a disutility for the clinical state and assumed this also applies to screen-detected individuals once their disease progressed to the point that it would have presented symptomatically in the absence of screening. This framework discounted the costs and effects on a discrete annual basis using a user-defined discount rate and discount year using the standard exponential form.

The cost-effectiveness outcomes

This study records the principal cost-effectiveness outcomes, including the discounted and undiscounted costs, LYs gained, QALYs gained, and the set of strategies on the efficient frontier and the ICERs between them. This model also records intermediate outcomes including the age entering health states, a disaggregation of costs, and the numbers of screens, individuals entering the preclinical state in disease history, true-positives, and false-positives, cancer deaths, and clinical cases. The intermediate outcomes also include the over-diagnosed cases, i.e., individuals that were screen-detected but in the absence of screening would not have presented symptomatically before their (other cause) death.

User interface

To make the framework accessible for those unfamiliar with R, all the parameters can be defined in Excel. An Excel template with VBA code embedded is prepared for saving parameter inputs and reading main outcomes. The inputs defined in Excel are saved in separate files which R then imports. R is the main programme to execute the model. Similarly, the main model outputs can be accessed by Excel or R, including cost-effectiveness tables with ICERs and cost-effectiveness plane. The additional results are saved in separate output files which can be read by R.

An R Shiny app offers an intuitive interface with which to adjust model inputs and observe the changes in outputs (Figure 3-3). The Shiny app does not conduct model runs itself but rather facilitates the dynamic inspection of previously calculated results. It permits sensitivity analysis for parameter values, adjustment of the cost-effectiveness

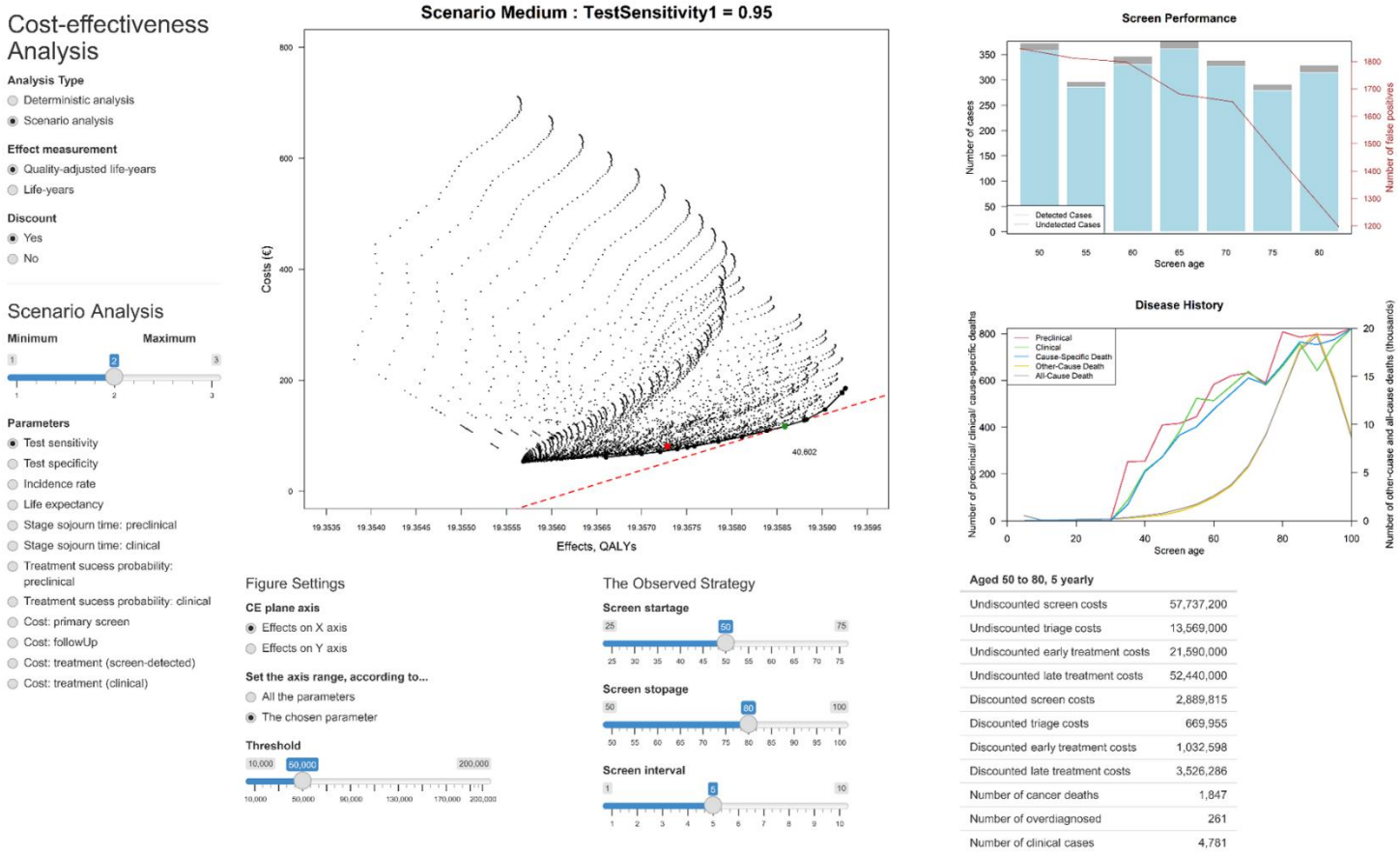
threshold and quick identification of strategies of policy interest. The Shiny app also plots the intermediate outcomes which can be useful if the user wishes to interrogate the association between strategy characteristics and cost-effectiveness.

The Shiny app allows the user to change the background settings, including analysis type (base-case vs scenario analysis), effect measurement (QALYs or LYs), results discounted or not, axis orientation and range. The user can choose the parameter they wish to vary, the parameter values (selecting from a predetermined three-point scale), a cost-effectiveness threshold value, and a specific screening strategy to observe (defined by the screening starting age, stop age, and interval). The user can also select the characteristics of a particular strategy of interest which is shown in the plot with the red marker. The intermediate outcomes for the observed strategy are also displayed, including the screen performance and the disease history.

Files in this framework

On GitHub, the user can find four files and two directories. Files include *ReadMe.txt*, *ExcelMaster.xls*, *Simulation.R*, and *ShinyApp_Plot.R*. Directories include *InputFiles* and *OutputFiles*. The content of *ReadMe.txt* also shows on the main GitHub page. The Excel template, *ExcelMaster.xls*, has a total of 21 spreadsheets, including model introduction, model inputs, and model outs. *Simulation.R* and *ShinyApp_Plot.R* are R files that can be inspected or edited with a plain text editor, such as Notepad++, in addition to R and RStudio. The folder of *OutputFiles* has additional folders inside, so downloading the complete packages is recommended.

Figure 3-3 Interface built in R Shiny app



The complete code for producing this Shiny app interface is provided on GitHub. This interface only plots the saved results from the main simulation model and does not perform real-time simulations. This interface gives users flexibility in choosing parameter values and figure settings. If "Deterministic analysis" is chosen for the analysis type, the choices in the section of scenario analysis are disabled. Under scenario analysis, numbers one to three respectively refer to the lowest value, the base-case value, and the highest value. Regarding the cost-effectiveness plane plotted in the middle of the interface, the green dot represents the optimal strategy identified with the specified threshold value in the Shiny app, and the red dot represents the observed strategy, defined by three variables: screening starting age, stop age, and interval. The other two diagrams and the table on the right-hand side are the intermediate outcomes.

Simulation.R

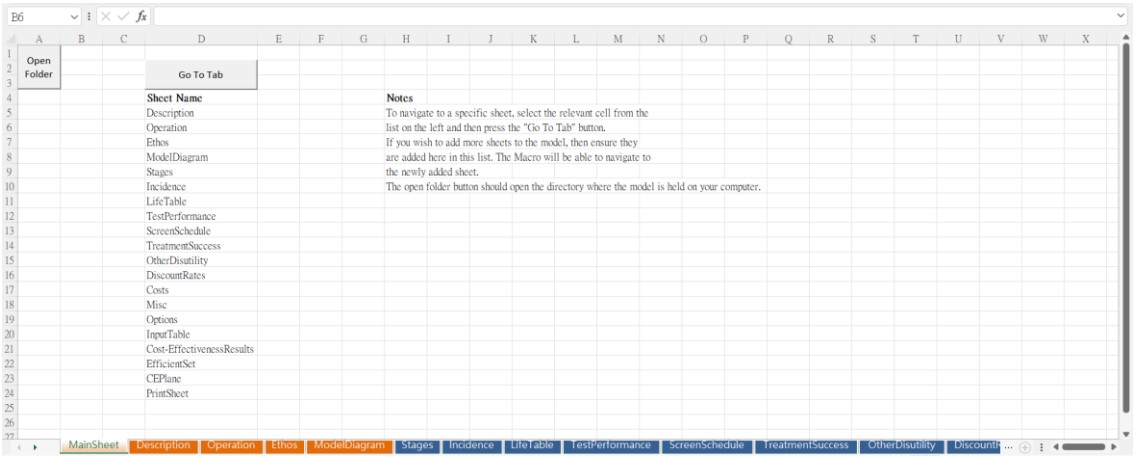
The *Simulation.R* file is the main file for running the model. This file contains the code for drawing input values from the input files. It contains the model code for the structure of the model and its operation. It also includes the code for producing possible screening interventions. The *Simulation.R* file contains code markup and is structured in accordance with the model structure and operation. The model outputs will be saved in the *OutputFiles* folder in whichever local directory you save these files.

ExcelMaster.xls

The ExcelMaster.xls file is an Excel file that allows users to define parameter input values, including the lower and upper values for the scenario analysis, and screening schedule details such as the frequency of screening and the age range without having to use R. This file is intended for use by those less familiar with R. The file uses VBA Macros to navigate through various portions of Excel and to write the input to files for the R model to then use. The macros are used by pressing the buttons within the Excel file. To ensure the macros can operate you may need to grant permission when opening the Excel file. Each tab within the ExcelMaster.xls file contains a brief description of the function of the sheet, including the definition of parameters and saving of input files for execution.

The model introduction has four spreadsheets, namely *MainSheet*, *Description*, *Operation,* and *Ethos*. The worksheet employs VBA macros to automate some basic operations. The VBA Macros are linked to the buttons called *Open Folder* and *Go To Tab*. Clicking *Open Folder* opens the working folder where this Excel is saved. *Go To Tab* helps navigate to the chosen spreadsheet by clicking one of the cells from cell D5 to cell D25 and then clicking the button *Go To Tab* in the *MainSheet*.

Figure 3-4 Spreadsheet in the provided Excel file (*Mainsheet*)

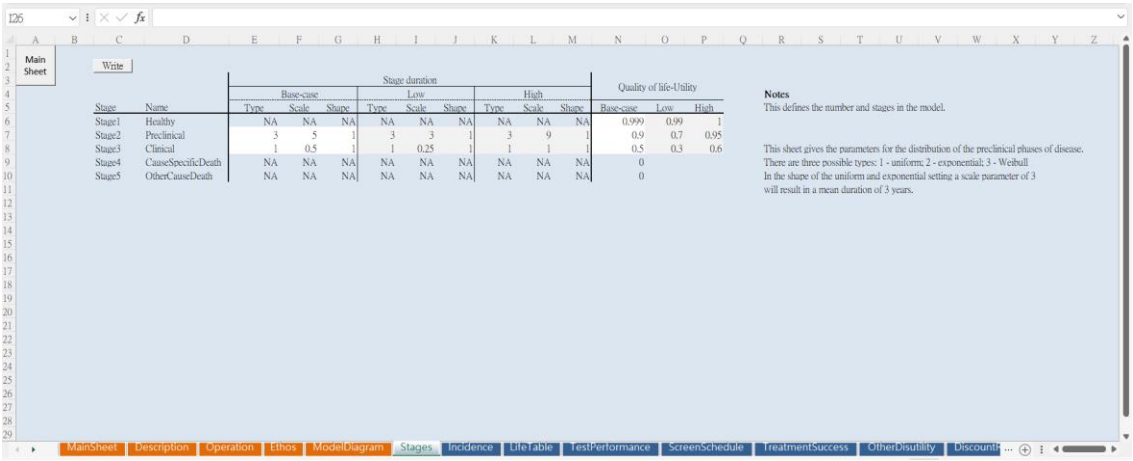


The Visual Basic for Application (VBA) linked with the buttons "Go To Tab" and "Open Folder" is accessible in Excel. The user can navigate to the desired spreadsheet by clicking on the intended sheet name and then clicking on the "Go To Tab" button.

The blue spreadsheets in the Excel file are for defining the parameters inputs for the simulation, including *Stages*, *Incidence*, *LifeTable*, *TestPerformance*, *ScreenSchedule*, *TreatmentSuccess*, *OtherDisutility*, *DiscountRates*, *Costs*, *Misc*, and *Options*. White cells are the parameter inputs for the base-case scenario and grey cells are for scenario analysis. All the blue spreadsheets have VBA Macros linked to the button *Main Sheet* and *Write*. Button *Main Sheet* navigates the Excel back to the first spreadsheet, *Mainsheet*, for easier managing a large Excel file with multiple spreadsheets. After revising the parameter input, press the *Write* button to update the data in the *InputFiles* folder.

Spreadsheet *Stages* in the Excel file defines the disease history, the sojourn time and the utility weight of each health state. There are three possible types of distributions for defining the sojourn time, using numbers 1-3 to respectively label the uniform, exponential, and Weibull distributions. This framework has the flexibility to add extra health states by inserting lines between Line 7 and Line 8. Notably, the sojourn time of the health state is defined by the incidence of the disease and no sojourn time is defined for other-cause deaths and cause-specific deaths because they are absorbing states.

Figure 3-5 Spreadsheet in the provided Excel file (*Stages*)

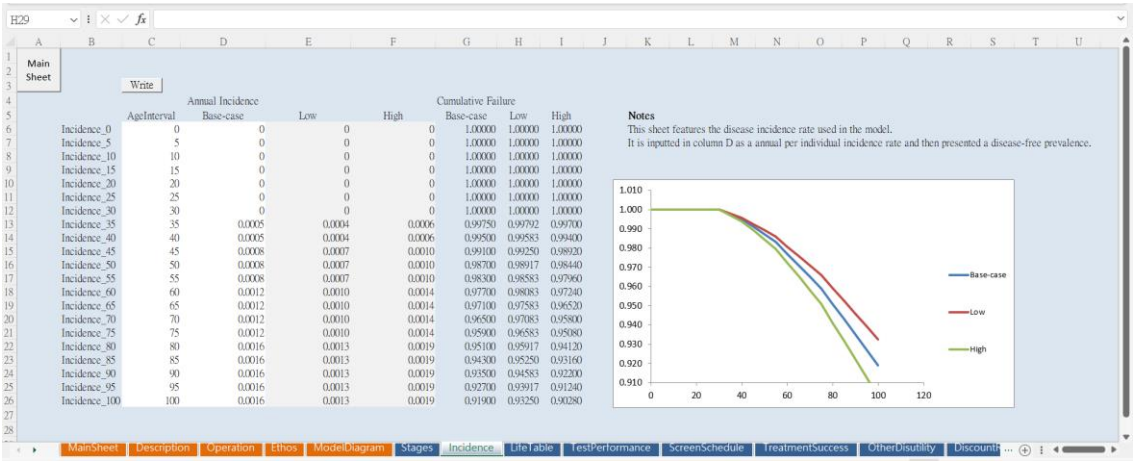


The Visual Basic for Application (VBA) linked with the buttons “Main Sheet” and “Write” is accessible in Excel. The user can navigate to another spreadsheet by clicking on the “Main Sheet” button. The white cells represent values for the base-case analysis, while the grey cells are for the scenario analysis. After making any changes in the white and grey cells, click on the "Write" button to save the updated parameter inputs.

Spreadsheet *Incidence* defines the incidence of the disease. The first column of white cells defines the age interval of the probability of incidence applied and the second column defines the corresponding annual incidence. For example, Figure 3-6 describes the base-case scenario that the studied cohort has a 0.05% chance of getting the disease between ages 30-35. Column B is the parameter name used in the R model. Column G is the cumulative incidence which means by the time of a specific age, the proportion of the studied cohort have the disease. Similarly, spreadsheet *LifeTable* defines the cumulative survival probability, meaning the proportion of the studied cohort died from other-cause before a specific age.

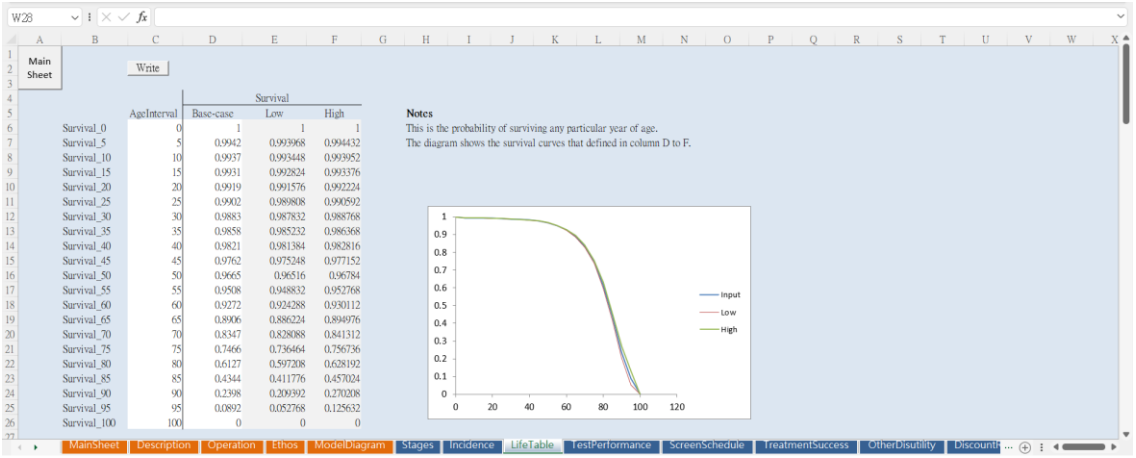
Regarding the screening, spreadsheet *TestPerformance* records the sensitivity, specificity, and dis-utility of screening tests. A total of five screening tests can be set up. Spreadsheet *ScreenSchedule* depicts the screening schedule in Figure 3-2. Further details can be found in Section 2.2.2. In default, only strategies with a single interval and a single screening test are considered, so the parameters related to the switches of test or interval are NA.

Figure 3-6 Spreadsheet in the provided Excel file (*Incidence*)



The Visual Basic for Application (VBA) linked with the buttons “Main Sheet” and “Write” is accessible in Excel. The user can navigate to another spreadsheet by clicking on the “Main Sheet” button. The white cells represent values for the base-case analysis, while the grey cells are for the scenario analysis. After making any changes in the white and grey cells, click on the "Write" button to save the updated parameter inputs. The variable names listed in column B are employed in the model to represent the annual incidence for subgroups divided by each five-year age bracket.

Figure 3-7 Spreadsheet in the provided Excel file (*LifeTable*)



The Visual Basic for Application (VBA) linked with the buttons “Main Sheet” and “Write” is accessible in Excel. The user can navigate to another spreadsheet by clicking on the “Main Sheet” button. The white cells represent values for the base-case analysis, while the grey cells are for the scenario analysis. After making any changes in the white and grey cells, click on the "Write" button to save the updated parameter inputs. The variable names listed in column B are utilized in the model to denote the survival rates for every five years of age.

For the effectiveness of treatment, spreadsheet *TreatmentSuccess* gives the probability of treatment success given the stage of detection, including clinical and preclinical probabilities. Spreadsheet *OtherDisutility* records the one-time dis-utilities of triage and treatment, respectively. Spreadsheet *Costs* lists the costs of primary screening, follow-up investigation of primary positives, treatment for screen-detected, and treatment for clinical-detected.

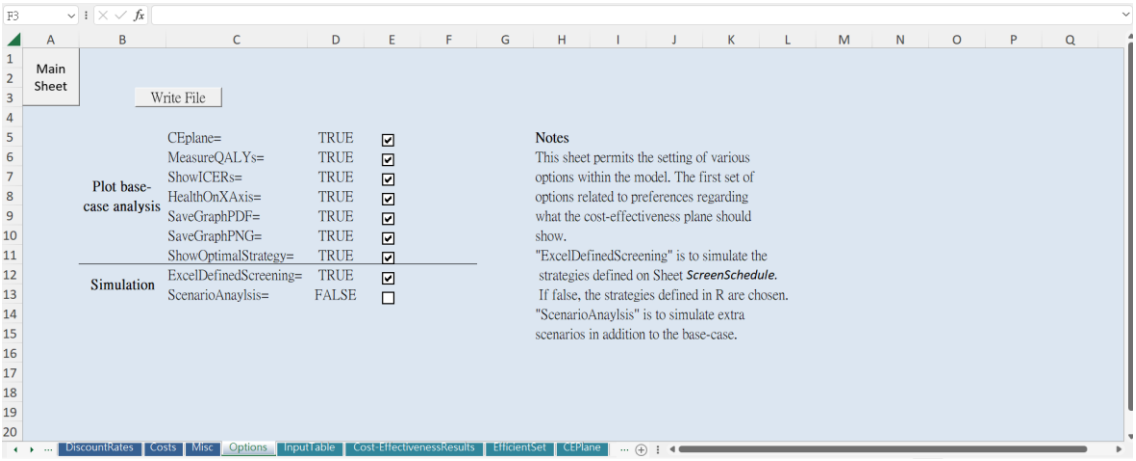
Spreadsheet *DiscountRates* characterises the discounting rates, respectively for costs and effectiveness. It also has the parameter *DiscountYear* to adjust the base year for calculation. The default of parameter *DisountYear* being zero means the simulation starts discounting since the first year of the simulation. Spreadsheet *Misc* details the sample size of the simulation and the decision threshold.

The administrative parameters are also available in the *Options* spreadsheet to adjust the settings for the presentation of results on the cost-effectiveness plane and to choose if simulate the strategies defined in Excel and execute the scenario analysis. If the blank in the same row is not ticked, it means that the screening strategies defined in R are adopted, instead of the screening strategies in spreadsheet *ScreenSchedule,* or that only run the base-case scenario. The scenario analysis here refers to the one-way sensitivity analysis (OWSA), in which the simulation only changes one parameter at a time.

Users can choose if a cost-effectiveness plane is plotted by parameter *CEplane*. Parameter *MeasureQALYs* controls which effectiveness measure, discounted QALys or LYs, is plotted on the cost-effectiveness plane. Parameter *ShowICERs* is for labelling the ICERs for the strategies on the efficient frontier. Parameter *HealthOnXAxis* is for the choice of plotting effectiveness on the X axis or on the Y axis. Parameter *ShowOptimalStrategy* is for applying the label of the optimal strategy on the cost-effectiveness plane. Parameter *SaveGraphPDF* and parameter *SaveGraphPNG* are to decide if to save the diagram to a format of .pdf or .png file.

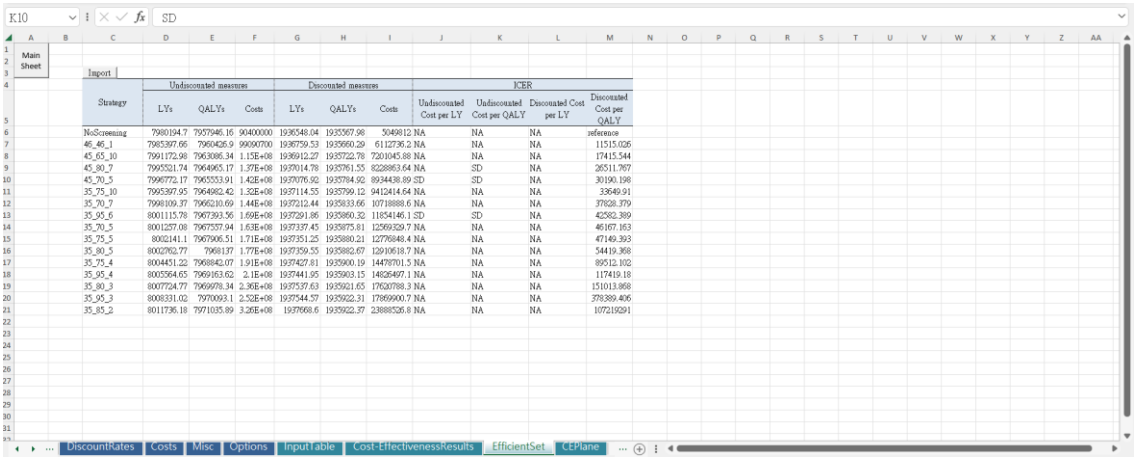
To inspect the updated results, press the “Import” or “Update” button to view the latest results. The cost-effectiveness results for all the strategies and the efficient set are recorded in separate spreadsheets, *Cost-EffectivenessResults* and *EfficientSet*. The effectiveness measure for choosing the efficient set depends on the parameter *MeasureQALYs* in the *Options* sheet. The cost-effectiveness plane is plotted in spreadsheet *CEPlane*. This Excel document also provides a table in spreadsheet *InputTable* for overviewing all the parameter inputs in the result section. The user should leave the last spreadsheet of this file, *PrintSheet*, blank for the operation of embedded VBA Macros.

Figure 3-8 Spreadsheet in the provided Excel file (*Options*)



The Visual Basic for Application (VBA) linked with the buttons “Main Sheet” and “Write” is accessible in Excel. The user can navigate to another spreadsheet by clicking on the “Main Sheet” button. After making any changes, click on the "Write File" button to save the updated administrative parameter inputs. Please refer to the main text of this thesis for the meaning of each administrative parameter.

Figure 3-9 Example of output estimates for efficient strategies in the Excel file (*EfficientSet*)



ICER, incremental cost-effectiveness ratio; LYs, life-years; QALYs, quality-adjusted life-years. After completing the simulation in R, click on the 'Import' button to access the updated results. This spreadsheet presents the numerical cost-effectiveness results of the strategies that lie on the efficient frontier, assuming discounted QALYs as the primary outcome for analysis. If users desire to access the results for the efficient frontier based on discounted LYs, please refer to the options spreadsheet to modify the settings.

ShinyApp_Plot.R

The ShinyApp_Plot.R file contains the code for running the package *Shiny* which provides a GUI within which to inspect the results of the simulation model and to move between alternative parameter sets in order to inspect the effect of different parameters on model outcomes. The Shiny app does not run the model in real-time but rather relies on previously generated output previously generated by the *Simulation.R* model.

In addition to the cost-effectiveness plane, the intermediate outcomes for a specific screening strategy can be accessed through Shiny. The table below lists the discounted and undiscounted intermediate outcomes and the number of cancer deaths, over-diagnosis, and clinical cases. The following plots show the screening performance and disease history of this screening strategy.

The following options are to customise the results displayed:

1. Choose the analysis type. Choosing deterministic analysis will plot the base-case scenario. Only after scenario analysis has been executed in *Simulation.R*, the results for scenario analysis can be plotted by Shiny.
2. Choose effect measurement, either QALys or LYs.
3. Choose whether to discount the results. The discount rate can be changed in *ExcelMaster.xls*.
4. Specify the observed scenario by moving the slider and the parameter of interest. The minimum setting on the slider is the lower parameter value. Maximum is the upper value. The middle scenario refers to the base-case scenario.
5. Choose which orientation of cost and effects to use on the cost-effectiveness plane. All the simulated screening strategies are plotted in this diagram.
6. Choose the axis range. The range can be large enough to show all the scenarios or the range can be suitable for showing the variance of the observed parameter that is specified in step 4.
7. Vary the cost-effectiveness threshold value by moving the threshold slider. The available range of the decision threshold is €10,000-€200,000 per unit of the effectiveness.
8. Specify the characteristics of observed strategy regarding the screening start age, stop age, and interval.

An example

To demonstrate the model, this study simulates 100,000 individuals with an illustrative parameter set. Assuming the screen interval remains fixed throughout the programme is an integer in the range of 1-10 years and screening start and stop ages between 25 and 100 we identified all possible screening strategies. This study also considered any one-off screening within this age range. In total, 8,161 screening strategies were simulated, including a no-screening strategy. The cost-effectiveness threshold in the example is €50,000 per QALY. Regarding the efficiency of the simulation, this study measures the execution time of one simulation with a computer having 3.2 GHz I7-8700 processor with 32 GB RAM.

To illustrate the relationship between parameter values and cost-effectiveness, the simulations were repeated over a range of alternative parameter values. The Excel file on GitHub lists all the parameter values used in this example. The studied parameters include screen cost, treatment cost, treatment effectiveness, test performance (test sensitivity and specificity), incidence and other-cause mortality, and the sojourn time of health states (preclinical and clinical).

This study employs comparative statics to show the efficient frontiers within the cost-effectiveness plane before and after a change in parameter values. The process of comparative statics involves comparing the results of the model with a change in one or more parameters while holding all else equal. It is instructive to view two sets of results: (i) those showing the absolute costs and health effects of all strategies including no-screening; and, (ii) those illustrating costs gained and health effects gained relative to the no-screening strategy.

This study observed the strategies on the efficient frontiers in each scenario and the direction of the efficient frontier moving on the cost-effectiveness plane as a parameter changes. This study separately describes the observations for the impact on cost and effect estimates. This study also demonstrated how the position of the efficient frontier changes suggests the change in the cost-effectiveness of screening.

Results

This section presents the results of an example using the pedagogical screening model developed by this thesis. This example simulated 8,161 strategies, including a no-screening strategy. These comprise screening intervals ranging from 1 to 10 years with start and stop ages ranging from 25 to 100. The one-off screening is also considered in this analysis. This section demonstrates the movement of the efficient frontier over a range of parameter values on the cost-effectiveness plane with the absolute estimates and the estimates relative to the no-screening.

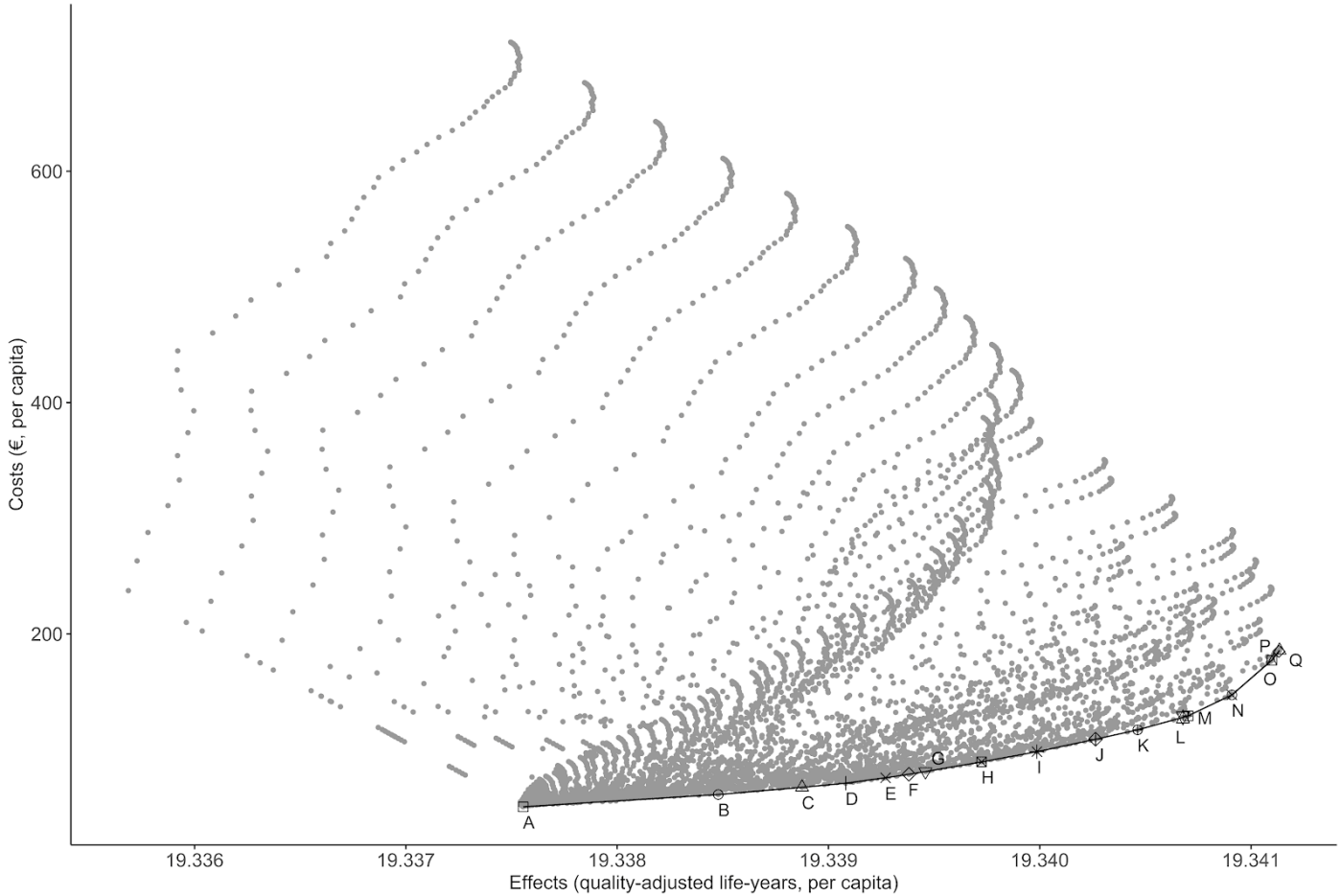
The execution time is 1.9 minutes for 250 screening strategies, and 1.45 hours for a complete simulation of all 8,161 strategies on a 3.2 GHz I7-8700 processor with 32 GB RAM.

Cost-effectiveness results

This study identified 17 strategies on the efficient frontier with a broad range of ICERs including a no-screening strategy in the base-case scenario (Table 3-1 and Figure 3-10). These strategies are net costly and effective, increasing with frequency. The most intensive strategies are more effective, but not cost-effective. The efficient frontier's shape reflects diminishing marginal returns of screening intensification. In this example, the optimally cost-effective strategy in the base-case is screening every 6 years from ages 35 to 77 with an ICER of €40,602 per QALY.

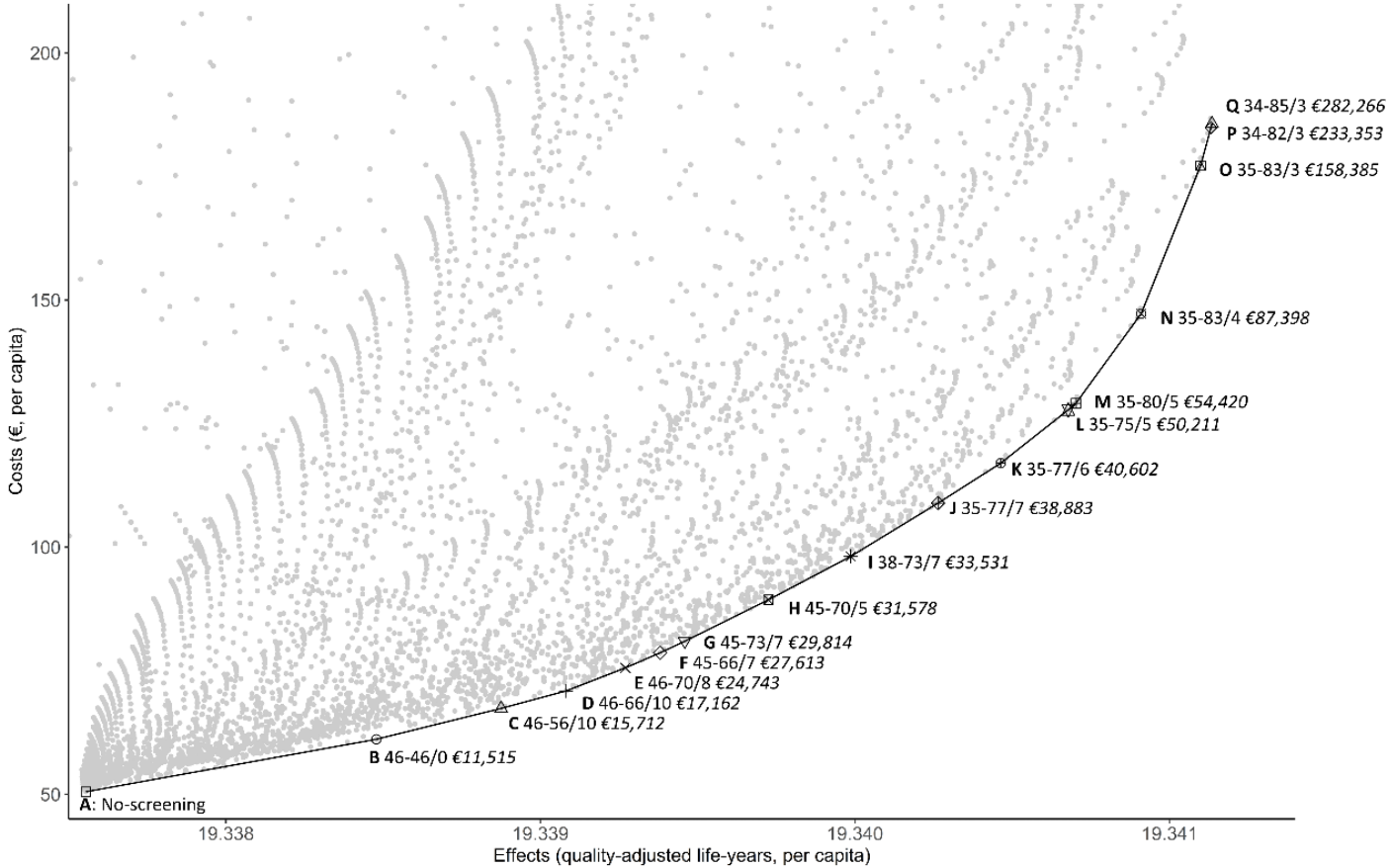
Figure 3-3 shows the Shiny app illustrating an example strategy of 5 yearly screening between ages 50 and 80. The results are plotted on the cost-effectiveness plane with the strategies forming the efficient frontier joined by the solid black line, the strategy with the optimal net health benefit shown with the green marker and the cost-effectiveness threshold with the red line. This strategy generates 4,781 clinical cases, 1,847 cancer deaths and 261 over-diagnosed cases.

Figure 3-10 The cost-effectiveness plane for the base-case scenario



This cost-effectiveness plane features 8,161 plotted strategies, including the no-screening scenario. These strategies encompass screening intervals ranging from 1 to 10 years, with start and stop ages varying from 25 to 100. Strategy A represents the no-screening scenario, which has the lowest cost estimate in this analysis. However, it may not have the lowest effectiveness estimate, as some screening strategies may result in more quality-of-life burdens than the benefits they offer. For details on the strategies located on the efficient frontier, please refer to Table 3-1.

Figure 3-11 The cost-effectiveness efficient frontier for the base-case scenario



Each strategy on the efficient frontier is labelled with the range of screening ages, interval, and its ICER estimate. Detailed results can be found in Table 3-1. These results represent a subset of 8,161 strategies, where only the strategies with higher or similar effectiveness to the no-screening scenario, as well as the strategies with lower or similar cost estimates to strategy Q on the efficient frontier, are plotted.

Table 3-1 The cost-effectiveness results for the base-case scenario

Strategy				Undiscounted outcomes			Discounted outcomes			ICER, €/QALY
Label	**Start age**	**Stop age**	**Interval**	**LYs**	**QALYs**	**Costs (€)**	**LYs**	**QALYs**	**Costs (€)**	
A	No-screening			79.801947	79.516654	904.00	19.365480	19.337555	50.498	-
B	46	46	0	79.853977	79.541461	990.91	19.367595	19.338478	61.127	11,515
C	46	56	10	79.886550	79.556584	1073.01	19.368548	19.338875	67.361	15,712
D	46	66	10	79.910897	79.567635	1146.67	19.369057	19.339082	70.908	17,162
E	46	70	8	79.930444	79.576316	1230.96	19.369545	19.339270	75.571	24,743
F	45	66	7	79.934988	79.578538	1239.68	19.369834	19.339381	78.626	27,613
G	45	73	7	79.947866	79.58399	1311.88	19.370050	19.339459	80.960	29,814
H	45	70	5	79.967722	79.592732	1418.12	19.370769	19.339725	89.344	31,578
I	38	73	7	79.973945	79.595844	1415.95	19.371485	19.339986	98.124	33,531
J	35	77	7	79.991348	79.603523	1501.67	19.372273	19.340264	108.922	38,883
K	35	77	6	80.005383	79.609435	1596.73	19.37285	19.340463	116.996	40,602
L	35	75	5	80.021411	79.616258	1707.34	19.373513	19.340678	127.768	50,211
M	35	80	5	80.027628	79.618563	1766.25	19.373595	19.340702	129.106	54,420
N	35	83	4	80.053901	79.628843	2021.69	19.374402	19.340910	147.236	87,398
O	35	83	3	80.080173	79.637804	2409.42	19.375412	19.341099	177.200	158,385
P	34	82	3	80.085183	79.640079	2436.57	19.375637	19.341132	184.887	233,353
Q	34	85	3	80.087298	79.640566	2485.51	19.375661	19.341135	185.699	282,266

ICER, incremental cost-effectiveness ratio; LYs, life-years; QALYs, quality-adjusted life-years. This table only shows the results for the strategies on the efficient frontier in the base-case analysis based on the estimates of discounted QALYs and costs. The cost and effectiveness estimates of other strategies can be generated using the files shared on GitHub.

Comparative statics

Figure 3-12 and Figure 3-13 demonstrate the efficient frontier over a range of parameter values. Notably, the strategies that comprise the frontier differ as parameters change.

(1) Screen cost

Changes in screening costs result in changes in the vertical plane only as these do not influence screening effectiveness. Lower screening costs makes screening more cost-effective, reducing the ICERs of all efficient strategies. The no-screening strategy is not influenced by changes in either the costs of primary screening or following triage, so the origin of the frontier remains static across scenarios. Both relative and absolute cost estimates decrease as screening cost decreases.

(2) Treatment cost

Treatment cost changes only influence overall costs without any change in effectiveness, so strategies simply move vertically in the cost-effectiveness plane. ICERs of screening decrease when the treatment costs decrease for the screen-detected or increase for symptomatic disease.

Changes to the cost of screen-detected disease do not influence the position of the no-screening strategy. Consequently, the absolute and relative costs of screening fall identically when the treatment cost for screen-detected disease falls. Conversely, varying treatment costs for symptomatic disease influences costs of both screening and no-screening strategies. While absolute costs increase, they increase the most for no-screening and by increasingly less as screening gets more effective. Therefore, a decline in the cost of treating symptomatically detected disease results in a fall in the cost of the no-screening scenario and the relative costs of screening rise and the cost-effectiveness of screening deteriorates.

(3) Treatment effectiveness

Changes in treatment effectiveness only result in changes in the horizontal plane. Changes to the effectiveness of screen-detected disease do not influence the position of the no-screening strategy, so the relative and absolute changes are identical. Conversely,

changes to the effectiveness of symptomatically-detected disease will influence the position of the no-screening scenario and the relative and absolute outcomes differ.

An improvement in treatment effectiveness of screen-detected disease shifts the outcomes to the right and all screening strategies become more cost-effective. A reduction in treatment effectiveness of symptomatically-detected disease shifts the frontier to the left in terms of the absolute estimates and shifts to the right in terms of estimates relative to no-screening and all strategies become more cost-effective.

If the effectiveness of early treatment falls to parity with that of late treatment then there is no advantage of screening and no-screening becomes the preferred strategy. There is a minimum difference in effectiveness between early and late stage treatment required for screening to ever be beneficial. This minimal difference is required, in part, to ensure that the advantages of screening at least outweigh the QoL losses imposed by screening itself.

(4) Test performance

Changes in test sensitivity or specificity influence screening effectiveness and costs, resulting in movement in both the horizontal and vertical plane. Improved test sensitivity or specificity improves the accuracy of screening results, and screening then becomes more effective, less costly and more cost-effective. As the no-screening strategy is not influenced by changes in test performance the absolute and relative outcomes are identical.

(5) Incidence and other-cause mortality

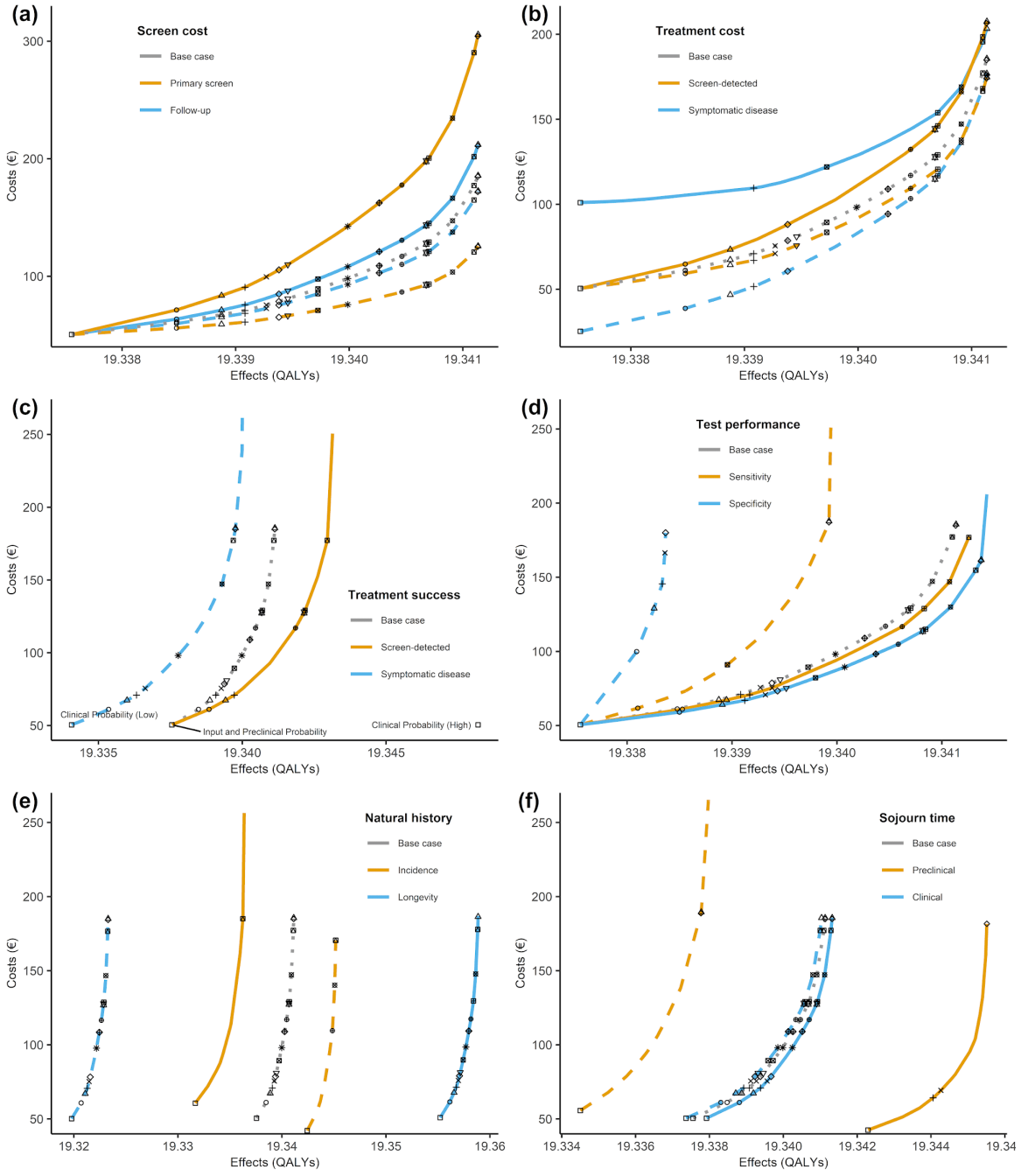
Varying disease incidence and other-cause mortality affect both cost and effectiveness of screening. Both parameters influence the cost and effectiveness estimates of the no-screening strategy, so the absolute and relative outcomes differ. In terms of absolute results, increased disease incidence leads to higher absolute costs and lower absolute effects for all strategies including no-screening. In terms of results relative to no-screening the converse is observed, as the relative costs of screening fall and the relative effects increase, meaning cost-effectiveness increases. Lengthening life expectancy (reducing other-cause mortality) increases both absolute and relative costs and health

effects. In this example, the relative outcomes indicate the efficient frontier moves to the right, indicating screening becomes more cost-effective.

(6) Sojourn time

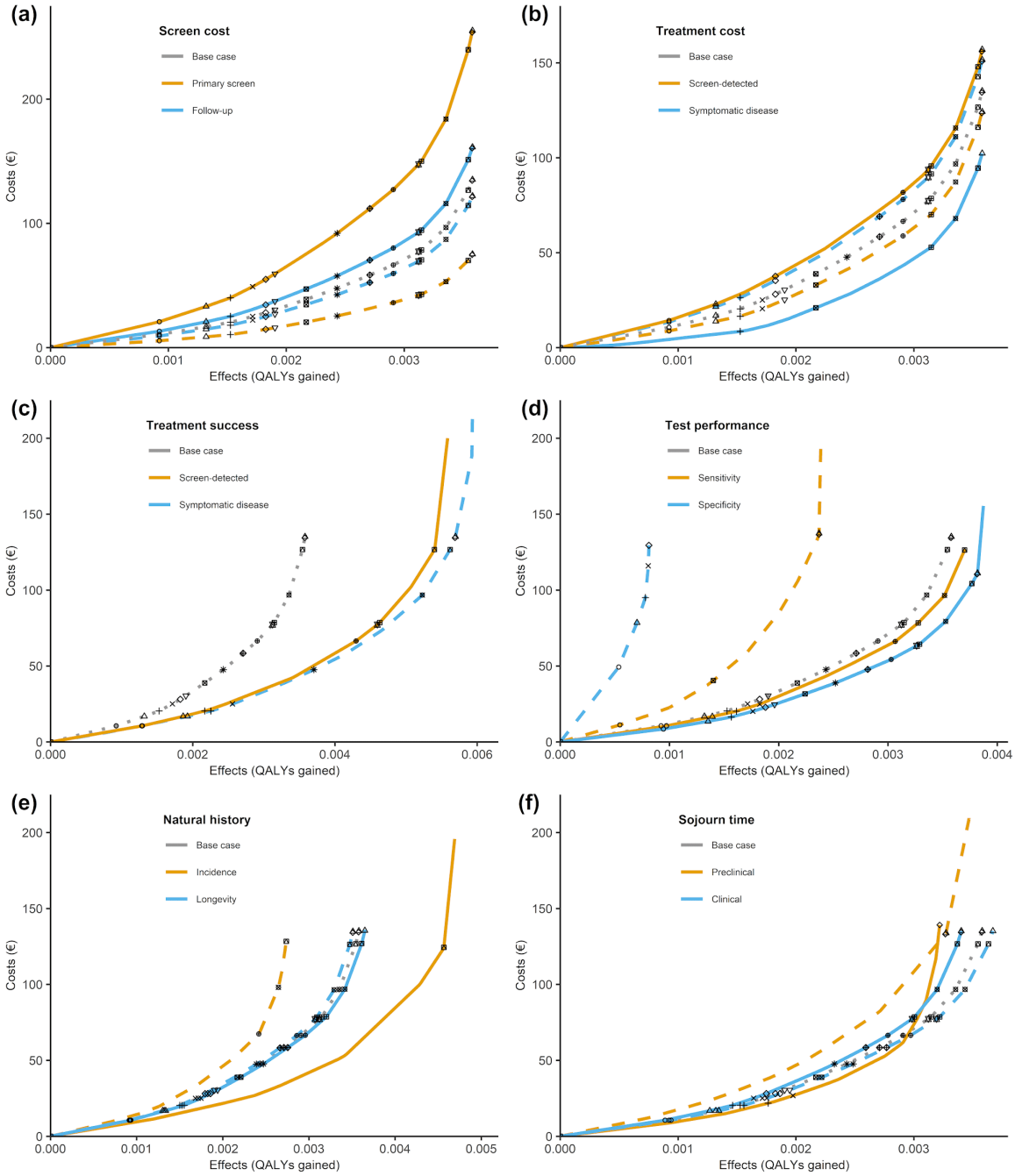
In absolute terms, a lengthening of preclinical or clinical sojourn time results in increased effectiveness. Absolute costs reduce with an increase in the preclinical sojourn time but remain unchanged with the clinical sojourn time. In relative terms, lengthening the preclinical sojourn time has an ambiguous effect on effectiveness and cost-effectiveness. Some low intensity strategies become relatively more effective, but higher intensity strategies become relatively less effective. Consequently, the ICERs for some strategies fall and rise for others. The relative outcomes for lengthening the clinical sojourn time are unambiguous as effects fall and costs remain unchanged, meaning the ICERs rise for all strategies as screening becomes less cost-effective.

Figure 3-12 The cost-effectiveness efficient frontier over a range of parameter values (absolute results)



QALYs, quality-adjusted life-years. Solid line represents the higher-value scenario and dashed lines are the low-value scenarios; the shape of the dots is corresponding to the strategies in Figure 3-11. In the scenarios of high treatment success for symptomatic disease and low for screen-detected, the no-screening strategy becomes the only strategy on the efficient frontier. Noted that the range of axis are different to better illustrate the shape change across the scenarios of each parameter sets. The grey dotted line represents the base-case scenario, where the estimates remain constant across all the diagrams.

Figure 3-13 The cost-effectiveness efficient frontier over a range of parameter values (results relative to no-screening)



QALYs, quality-adjusted life-years. Solid line represents the higher-value scenario and dashed lines are the low-value scenarios; the shape of the dots is corresponding to the strategies in Figure 3-11. In the scenarios of high treatment success for symptomatic disease and low for screen-detected, the no-screening strategy becomes the only strategy on the efficient frontier. Noted that the range of axis are different to better illustrate the shape change across the scenarios of each parameter set. The grey dotted line represents the base-case scenario, where the estimates remain constant across all the diagrams.

Discussion

Brief findings

This study presents a simplified screening CEA microsimulation model for teaching and research purposes. As an initial application, this study presents an assessment over a large range of strategies and conducts comparative statics to illustrate the influence of parameters on cost-effectiveness. The results of this study illustrate the relevance of considering both absolute costs and effects and those relative to no-screening. This distinction between absolute and relative outcomes is useful when seeking to demonstrate the intuition behind the observed results. The analysis of this study conveys the intuition of the relationship between parameter values and outcomes, informing the process of model validation.

The first study of this thesis found methodological shortcomings in CEAs of CRC screening. This, may in part, be due to the fact that there is no open-source model designed for demonstration of methods in the screening CEAs. Established research groups usually do not share their CEA models of screening completely, e.g., MISCAN (Goede et al., 2013; Van Der Meulen et al., 2018; Wilschut, Habbema, et al., 2011; Wilschut, Hol, et al., 2011). This is understandable partly because the applied models are very complex and are typically designed for specific research questions. These applied models might be easily misused if applied by those with the advanced technical understanding required to operate them. This may partly explain why their authors may be reluctant to share them publicly. This unwillingness to share, however, is counter to the ethos of transparency that is supposed to be adopted in academic research. Accordingly, this study addresses this issue by offering an open-source model that provides a neutral platform for modelling groups and individual researchers to demonstrate teaching and methods issues without having to share their own models.

This framework is accessible and editable by all as the complete model code and variable inputs are specified and provided online. Users are able to apply and extend the model without concerns of copyright infringement. This complete access contributes to research transparency and facilitates sharing knowledge of simulation methodology. As such, this model is intended as a public good and it is hoped that dissemination of this study will benefit the field of CEA in disease screening.

Advantages of this model

An important advantage of this model is its simplicity and speed. Compared to specialized commercial software or spreadsheet applications such as Microsoft Excel, the non-proprietorial nature of R permits an accessible, transparent, and adaptable model platform (Frederix et al., 2013; Incerti et al., 2019). Because R is free of charge, R is an accessible platform for low resource settings, including countries with modest per capita incomes or research and teaching settings without large budgets for purchasing software licenses. R is increasingly adopted as the modelling tool with the support of a large range of open-sources and well-documented packages and functions (Jalal et al., 2017). Importantly, models written in R are now accepted by NICE (NICE, 2022). There is an initiative using R as a health economic modelling tool and some open-source models are published for therapeutics simulation (Green et al., 2022; Sullivan et al., 2016). These models, however, are too complex for teaching purposes. In addition, those published tutorials for modelling in R are not applied to screening and do not employ DES (Frederix et al., 2013; Krijkamp et al., 2018; Williams et al., 2017). As such, this model offers a novel contribution to the growing R in CEA literature.

Coding transparency is another important advantage of this model. This model is only written in one single file and has self-defined functions in the same file, instead of saving each function in independent files. Saving the code in one single file helps readers access the complete code and improve the transparency of the model. The model is all coded in base R, without any application of C++ or packages. The only exception is the use of the Shiny package to build an advanced user interface for plotting the results, though this is not a must as the use of Excel and base R can also view the results. This model has extensive documentation for assisting users in understanding how to use this framework and a line-by-line markup.

Teaching tool

As a teaching tool, this model is intended for two groups. First, it can serve as a resource for students who want to understand the principles of economic evaluation regarding screening interventions. The intuitive interfaces of Excel and Shiny ensure that students do not need to understand R programming as they are able to explore alternative screening policies under different scenarios and threshold values without having to

operate or modify R code. The simulation model is deliberately written without dependency on any previous packages, removing the need for the user to know how to install packages (the sole exception being the installation of the Shiny package). The Shiny app example of this study offers a convenient interface for the examination of changes to parameter values on cost-effectiveness estimates. Secondly, this model serves as a resource for more advanced students intending to learn DES programming in R. This model provides a starting point for extension to other implementations, either methodological or applied.

This simplified model is suitable for the purpose of demonstrating the relationships between key parameter values and cost-effectiveness. Its simplified nature makes demonstrating face validity straightforward. While this simplified model can help modellers develop their understanding of screening, any specific modelling application requires independent, context-specific demonstration of validation. As such, any extension of this model might require a renewed exercise in face validity depending on how extensive the changes are.

This model deliberately employs a high degree of abstraction to make it accessible and efficient. Although this simulation only has five stages, it is sufficient to demonstrate the fundamentals of screening cost-effectiveness. As an abstracted model it is not suitable for solving applied research questions regarding specific prevention programmes. Rather, it is intended as offering a basis for addressing methods research questions. This model is able to demonstrate the strategy omission in the CEAs of screening which is addressed in the first study of this thesis. This model with the simplified structure can manage to simulate an extensive range of screening strategies with different characteristics that the applied models can barely do.

This pedagogical model can serve as platform to illustrative other methodological issues. It has the potential to demonstrate alternative forms of risk stratification, highlight distinctions between simulations of single and multiple birth cohorts, and illustrate the consequences of omitting screening strategies. Some CEAs examine alternative risk groups through separate analyses, while others employ a singular analysis to evaluate eligibility thresholds, assuming a uniform strategy for all screened individuals (Fabbro et al., 2022). Our model can effectively showcase the discrepancy in ICER estimates

between these two approaches by adjusting disease incidence or modifying the sojourn time of health states for different risk groups. Similarly, in CEA simulations involving single and multiple birth cohorts, our pedagogical model can illustrate the differences in ICER estimates by conducting multiple simulations with parameter specifications tailored to each respective cohort. Demonstrating the consequences of omitting screening strategies can be accomplished by intentionally excluding specific strategies considered in the analysis. The simplicity of the pedagogical model enhances its value in illustrating the methodology applied in CEAs without introducing unnecessary complexities.

Absolute vs relative outcomes

Model outputs can be considered in either or both absolute and relative terms. Absolute results are the absolute lifetime estimates of costs and effects without deduction of those of the no screening scenario. Relative results are those that are reported net of the no screening scenario, thus they report the difference vs the situation without screening. This distinction can be quite important when forming an intuitive understanding of the cost-effectiveness estimates and how they vary with parameter changes. This study has deliberately and explicitly presented both absolute and relative results in order to demonstrate their relevance to an intuitive understanding of how cost-effectiveness varies.

In many cases the absolute and relative results vary in the same way with a change in parameter values. For example, when test sensitivity increases both the absolute and relative effectiveness of screening increase. In other cases, the changes do not vary in the same way. For example, an improvement in treatment effectiveness for symptomatically detected disease increases the absolute effectiveness of all screening strategies but reduces their relative effectiveness. Where the values vary in opposite directions the intuition behind parameter changes may not be obvious. Accordingly, both forms of presentation are beneficial when validating a model. The following paragraphs separately give details to the absolute and relative results plotted on the cost-effectiveness plane.

It seems that absolute results may be those that are the most intuitive to interpret as they directly relate to the expected costs and effects of screening outcomes before netting out the effects relative to a no screening scenario. For example, when the screening cost

increases, the absolute cost estimates are higher for those screening more frequently which is the product of the unit cost and the number of resources consumed. Hence, the efficient frontier moves up on the cost-effectiveness plane in the example.

As an aside, absolute results do face constraints in terms of how comparable they may be between studies if analysts employ different start ages for simulation, run the models over different time horizons or if populations have different life expectancies. Though these considerations of comparisons between models is less relevant in the context of understanding the influence of parameter values on costs and effects.

Relative cost-effectiveness results refer to the presentation of costs and effects of each screening strategy relative to the no screening scenario. Such results make for easy calculation of ACERs by dividing the net costs by net effects. The relative results are those that are most relevant when we wish to consider the impact of a change in parameters on cost-effectiveness. For example, as treatment cost increases for later stage disease (typically symptomatically presenting cases), the absolute cost estimate increases for all the strategies, including the no-screening scenario. Because low-intensity screening strategies have more symptomatically-detected cases they have a higher increase in cost estimates, in contrast there will be a smaller increase among high-intensity screening strategies, which have relatively fewer symptomatically-detected cases. Hence, the efficient frontiers are observed moving down on the cost-effectiveness plane, which means screening becomes more cost-effective.

Understanding the difference between the absolute and relative results is important for understanding how results vary according to screening intensity. In the example described in the previous paragraph the absolute changes are larger for the low intensity strategies than for the high intensity strategies. This difference leads to non-equiproportionate changes in the relative results and this leads to a change in the efficient frontier, which becomes shallower, indicating that screening becomes more cost-effective across all strategies.

The example of an increase in the cost of late state treatment results is like many in that a change in the parameter value leads to a common direction of change for all strategies in the cost-effectiveness plane, resulting in unambiguous changes in cost-effectiveness.

The degree of change is associated with screening frequency in many cases. This degree of change, however, sometimes is most pronounced for screening at a middling frequency. This is observed when the test sensitivity and the preclinical duration are varied. In such cases there is no unambiguous change in the ICERs along the efficient frontier and the degree of convexity of the threshold can change. In such ambiguous cases it can be difficult to anticipate what the net effect of a parameter change on cost-effectiveness will be.

This study suggests CEAs should be careful when reporting and interpreting absolute and relative results. Both forms of reporting have been observed in the CEAs of screening (more details can be found in the discussion of the first study in this thesis). This thesis, however, does not find any previous literature that has compared both forms of reporting or discussed their relevance for interpreting cost-effectiveness evidence.

The cost-effectiveness outcomes of this study are for illustrative purposes, so this study avoids dwelling on the specific numerical outcomes and rather discusses the qualitative relationship between parameters and result cost and effects estimates and resulting ICERs. The specific numeric outputs for the model can be found on the GitHub repository. The complete cost-effectiveness outcomes and intermediate outcomes can be produced by running the default model with input values provided on GitHub.

Limitations

Naturally the model presented in this study has limitations. At a minimum, users must at least be able to install and run R. If they wish to use the Shiny application then they will need to install the "Shiny" package. The deliberate avoidance of packages elsewhere in the model results in minor imperfections in the presentation of overlapping ICERs within the cost-effectiveness plane.

The transparency of this model relies on the endurance of R and GitHub as this R model is shared via GitHub. Although the model framework can also be requested by contacting the authors, GitHub is a depository that largely improves all the transparency and version controls. Consequently, we chose to share this model via GitHub and this study naturally has drawbacks from using GitHub. For instance, the workflow of version controls can be

influenced if GitHub experiences issues or undergoes policy changes. Additionally, GitHub has limitations on file sizes and total storage, which can pose constraints when managing substantial files or extensive projects. Furthermore, as a public platform, GitHub faces security concerns, including the potential risk of accidentally exposing sensitive information. It's important to note that the aforementioned points are just a subset of the various drawbacks associated with its use.

A consequence of the degree of abstraction adopted in the model is that its structure and parameter values are merely notional and do not correspond directly to any specific disease. For example, a one-off treatment cost in this model cannot illustrate the treatment cost correlated with the severity of the disease or the length of hospital stay. The distinction between screen-detected and symptomatically detected treatment is a simplistic representation of early and late stage therapy. The model also does not include palliative care costs and death-related expenses.

This study did not demonstrate the external validity of the model, as the model was not specifically designed for a particular disease or intervention. The empirical observed data, such as disease incidence by age, cannot be directly compared to the presented incidence in this study. Furthermore, the practical application of the model's findings regarding the impact of parameters on the cost and effectiveness estimates of strategies is not discussed in this study as it is hampered by reporting constraints of OWSA in applied CEAs. Most CEAs only briefly reported the results of OWSA by presenting a tornado plot (McLeod et al., 2017; Ngan et al., 2022; van Rossum et al., 2011). The absence of a comprehensive distribution of cost-effectiveness outcomes for screening strategies impedes our ability to assess external validity. This limitation arises from the hypothetical nature of the model and the absence of detailed results documented in the literature.

As the authors (YSL & JOM) are not programmers, this model has not been through code optimisation. This programme is conducted in the way the authors think it is the most easily understood as this is designed for teaching. A practical limitation of this model is that although it achieves a fast runtime it will not retain such speeds when extended to the multiple health states and complex screening and triage algorithms required in applied analyses. The model code can be improved by sharing the model and improving its

transparency. Further adaptations might require integration with C++. Another limitation is that this initial demonstration does not explore parameter uncertainty. Adding probabilistic sensitivity analyses or exploring multivariate adjustment of input parameters are obvious future extensions.

Future research

This study identifies some research directions that might be fruitful for future studies. First, considering R models have a lot of advantages, more open-source R models are desired. It is hoped that sharing this model and its code will encourage more researchers to build models in R.

Second, the teaching model described in this thesis is an outcome of a trade-off between simplicity and representability, so this thesis hopes future research can extend this model to be more efficient or more representable. Possible technical extensions include applying existing packages, integration with C++, and parallel execution of R code with multiple computing processor cores or CPUs. In addition to the suggestions in the limitations section, future research can extend the model to include MCDA. For instance, the IVI-RA model, an open-source model, offers a user-interactive interface with a separate tab dedicated to showcasing the MCDA framework (Incerti, Curtis, et al., 2019). This model demonstrates how varying the weights of other attributes in decision-making impacts the value of the comparators (Incerti, Curtis, et al., 2019). As this thesis focuses on the cost-effectiveness of screening, this thesis does not consider other aspects that might influence the outcomes of decision-making.

Third, the association between the cost-effectiveness outcomes and the characteristics of screening strategies is still unclear. Future research can explore their relationship which may help understand the screening policy and its potential impact in terms of cost-effectiveness. Employing the distinction made here between absolute and relative results will support transparency in such work.

Fourth, as mentioned regarding the recommendations stemming from the systematic review, there is a need for screening-specific CEA guidelines. The simple model can support this process by providing a shareable basis for generating illustrative outcomes that demonstrate how such methods guidelines should be applied.

4. STUDY 3 AN IDEAL EFFICIENT FRONTIER OF SCREENING

The study applies the framework proposed in the second study to enhance our understanding of appropriate ICER generation and to address problems related to strategy omission, as identified in the first study. This study examines the distribution of strategies with constant characteristics of screening schedules, such as fixed screening start age, stop age, or the screening interval. Additionally, this study expands on the results from the second study by investigating the impact of parameter changes on the convexity of the efficient frontier. The study particularly emphasizes the significance of the policy-relevant portion of the efficient frontier, as it plays a crucial role in the final decision-making process. Ultimately, this study proposes an iterative method to identify relevant cost-effective strategies, offering CEA analysts a straightforward guide to self-examine if they have omitted critical strategies by observing their cost-effectiveness results.

Methods

Model description

This study applies the simple model described in the second study of this thesis and uses the same structure and assumptions. This study applied this individual-level DES framework of screening for an illustrative purpose to demonstrate the scenarios might imply strategy omissions. This framework has five health states, namely healthy, preclinical, clinical, other-cause death, and cause-specific death. A total of 100,000 individuals is simulated with an illustrative parameter set. The parameter values can be found in the Excel file provided by the second study of this thesis. All the simulations and diagrams are completed in R.

The main components of defining screening strategies are the testing modalities and the screening schedules. The primary source of strategy variation considered in this study is the timing of screening moments and this study regarded the characteristics of screening modalities, including test sensitivity, test specificity, and disutility, as model parameters. Same as the second study, the third study applied R to produce all the possible screening strategies with screening start and stop ages ranging between 25 to 100 years with a single interval of 1-10 years, assuming the screening interval remains fixed throughout the screening programme. The one-off screening at any age between 25-100 is also included.

A specific screening modality is assumed for all the screening services and is not varied during the course of the screening programme. In total, this example simulated 8,161 strategies, including the no-screening strategy.

Presentation tool

This study plotted the results on the cost-effectiveness plane, which offers a transparent way to compare the screening strategies and depict the impact of varying the cost-effectiveness threshold and other parameters of interest. Although other methods have been discussed for the presentation of the variation in cost-effectiveness outcomes over wide parameter inputs (Wolff et al., 2022), only the cost-effectiveness plane discloses the cost and effectiveness estimates of each simulation and links the results with the parameter inputs on the diagram. Accordingly, this thesis decided to apply the cost-effectiveness plane to plot the cost-effectiveness outcomes of interventions. A more detailed introduction to the cost-effectiveness plane can be found in the introduction chapter of this thesis.

Illustration of missing strategies

This study considered how the decision-making threshold relates to the cost-effectiveness results, in addition to the parameter uncertainty and heterogeneity of strategies. This study investigated the examples that might imply the omission of the optimal screening strategy from three features, including the characteristics of screening schedules, the decision-making with the efficient frontier, and the parameter values in the model. With the example, this study is able to give suggestions on how to expand the choice of strategies.

Heterogeneity of strategies

This study plotted all the strategies on the cost-effectiveness plane and observed their distribution based on screening starting age, stop age, and screening interval. In order to differentiate the individual impact of these three variables on cost-effectiveness, this study first identified strategies with the same two variables and then linked them together on the cost-effectiveness plane. For example, to inspect the starting age, this study linked strategies with the same stop age and interval, but different starting ages.

To better illustrate the marginal effect of an additional screen, this study separately plotted the cost-effectiveness planes from a subset of strategies. In the case of observing the effects of different starting ages, this study simulated screening strategies that stop at age 90 and start at age 30, 40, 50, 60, 70, and 80. Similarly, in another case of observing stop age, this study had strategies start at age 30 and stop every 10 years. For the screening interval, this study only plotted the strategies with screening intervals which can be perfectly divided by the defined age range.

Efficient frontier

Scenario analysis produces different shapes of efficient frontiers which influences the cost-effectiveness of screening. To illustrate the shape of efficient frontier and the decision-making, this study deliberately chose some parameters to implement OWSA, including lower test specificity (75%) and lower primary screen cost (€10). This study demonstrated the identification of optimal strategies under the assumed decision threshold between €20,000 and €30,000 per QALY. This study described the characteristics of an ideal efficient frontier and illustrated the importance of the policy-relevant portion of the efficient frontier with the examples.

Parameter values

The example of the second study in this thesis shows the cost-effectiveness results of all the key parameters, including screening cost, triage cost, treatment effectiveness, test performance, disease incidence, life expectancy, and the sojourn time of preclinical and clinical states. This study extended the second study by focusing on the convexity of efficient frontier and its impact on ICERs of screening strategies. This study implemented a OWSA of all the parameters and observed how the parameter changes the shape of the efficient frontier.

As an illustrative purpose, this study did not numerically measure the convexity of efficient frontiers. This study defines the frontier gets more convex as the ICERs of low-intensity screening on the efficient frontier become lower (frontier looks gentler) and the ICERs of high-intensity screening on the frontier get higher (frontier looks steeper).

For the parameters which are observed changing the convexity of the efficient frontier, this study further analysed their marginal change in effects and costs among strategies with different intensities. The strategies are chosen from the efficient set of the base-case scenario. A total of five levels of each parameter are simulated to observe the convexity of the efficient frontier. The change in effects or costs of a specific strategy between scenarios is calculated by the difference in the estimates before and after changing to the next level divided by the estimate before the change. This study also used the example to demonstrate the issue of strategy omission when there is a change in the convexity of the efficient frontier with parameter changes.

Possible solution

This thesis tested if an iterative method is able to solve the problem of comparator omission commonly observed in the CEAs of screening. The proposed method aligns with the recommendation from Drummond et al. (2015). Drummond et al. (2015) suggest not narrowing the decision problem but using some methods to exclude those who have a lower chance to be cost-effective. In practice, without at least some preliminary results, it is hard to anticipate which strategies might prove optimally cost-effective and most relevant for simulation, especially as there are multiple interacting factors impacting the cost-effectiveness outcomes.

The demonstration of complete strategy comparison guides appropriate screening strategy selection through an iterative approach which is more pragmatic in applied CEAs. The above examples in this study explained how to expand the strategies more efficiently if CEAs have had partial cost-effectiveness results for some strategies. Future CEAs can run a small selection of screening schedules and observe how their cost-effectiveness estimates are distributed on the cost-effectiveness plane. This observation can provide insights regarding how screening characteristics could be varied to yield more cost-effective strategies. The simulation can then be rerun with the expanded set of strategies. Furthermore, we can narrow the choice of screening strategies by focusing on the strategies with ICERs closer to the decision threshold. A total of five steps are identified.

First, CEAs of screening should define the scope of the study and identify all the possible strategies within this scope. These strategies should have variations in screening

modalities and screening schedules (age ranges and intervals). In this step, we consider as many strategies as possible without restricting the choice to resource constraints within the health system or conventional assumptions about what strategies are feasible and appropriate. Screening schedules are recommended to be defined systematically, separately setting up the range of starting ages, stop ages, and intervals. If different screening modalities are considered, then the same schedules should be considered within the given modalities to ensure they are assessed on a like for like basis regarding the accompanying screening schedules. Similarly, the analysis may need to consider alternative combinations of primary screening tests and triage algorithms, which again should be done on the basis of consistent variation of the screening schedules.

Second, from the broad range of strategies identified within the first step, select and simulate a manageable number of strategies. The selected screening characteristics should be widely and evenly distributed across the range of potential strategies. The number of chosen strategies depends on the simulation constraints of the analyst's computer, such as the size of memory capacity and average simulation time. Notably, the application of fixed random seeds to produce constant disease history or screening performance can help produce reasonable and comparable results across screening strategies with a smaller number of entities.

Third, in order to rule out the strategies with a low probability of being on the efficient frontier, the analyst observes the patterns of strategies on the cost-effectiveness plane. The cost-effectiveness plane can help analysts visually judge which characteristics tend to have higher effectiveness and lower costs. When discussing the patterns, linking the strategies with a common factor presents the impact of that factor on the cost and effectiveness estimates. From this observed pattern, the analyst can begin to infer what strategies are more likely to form the efficient frontier. Notably, we should include a broader range of these variables because the ICER estimate of a specific strategy is probably lower or this strategy might be dominated if we simulate more strategies. Generally, the range of strategies should be chosen when they are close enough to the frontier, not necessarily on the frontier with our prediction.

Fourth, simulate another set of screening strategies that are estimated to have a higher chance to be on the efficient frontier based on the observed pattern in the third step. This

step should apply the same seed as the second step to reduce stochastic noise. The simulation results should be saved in the format of the initial results, as it is easier to summarise all the simulations. The compartments of results should be saved separately, e.g., discounted and undiscounted effectiveness estimates. The results should include enough figures, especially as the difference in effectiveness is usually minor. The redundant results should be removed. For example, the results of the no-screening scenario may produce more than once.

Fifth, repeat the third and the fourth step until the strategies with characteristics of being a high chance to be cost-effective have been exhausted or other pragmatic considerations. The ideal decision range of the efficient frontier on the cost-effectiveness plane is a smooth curve in which the ICERs of strategies increase slowly.

This study provided an example using this iterative method to identify more strategies having a higher chance to be cost-effective. This study assesses if the strategies identified from the primary outcomes are more cost-effective by plotting them on the cost-effectiveness plane. This study then compared the efficient frontier from the iterative approach with the frontier from the full simulation of 8,161 strategies. The decision threshold is assumed between €20,000 and €30,000 per QALY.

Results

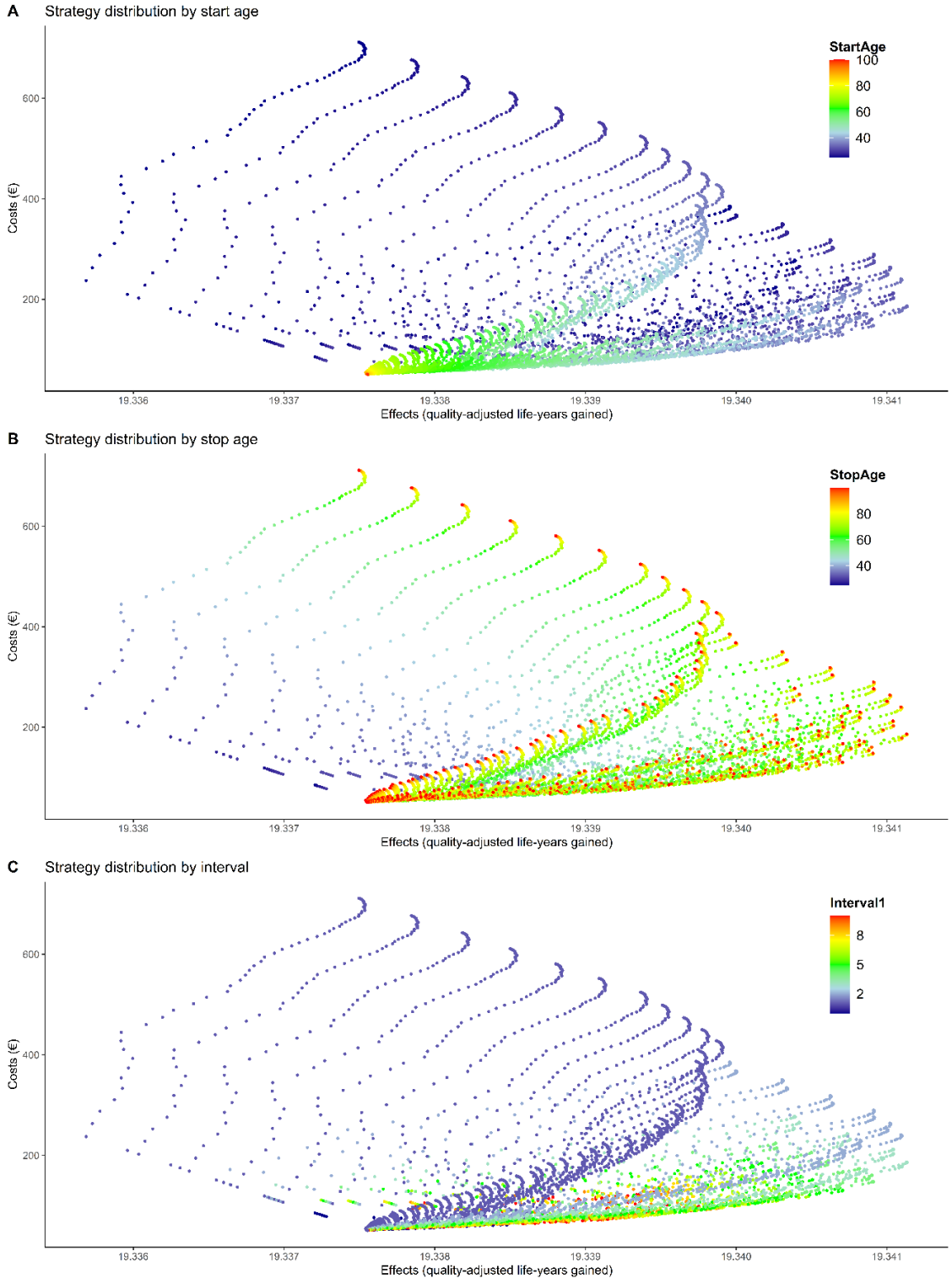
The third study takes a broader view of strategy choice within CEA, considering which portions of the efficient frontier should be most relevant to decision-makers, how this relates to the convexity of the efficient frontier and demonstrates how the problems of relevant strategy omission can be avoided. First, screening strategies are plotted within the cost-effectiveness plane according to the characteristics of the screening schedules and we examine how the positions of the estimates vary with the screening schedule characteristics. Second, the policy-relevant portion of the efficient frontier will be illustrated with an example. Third, this study identified the parameters that influence the convexity of the efficient frontier. This section explained how these three features are relevant to decision-making. Finally, an example is provided to explain how to expand the selection of strategies to avoid relevant strategy omission in the CEAs of screening.

General distribution patterns

This study revealed repeated similar patterns on the cost-effectiveness plane formed by the strategies with the same characteristics of screening schedules, including screening starting age, stop age, and interval (Figure 4-1). Screening strategies starting late in life, i.e., those over age 80, are clearly not cost-effective as their effects are very low, close to no-screening. Similarly, strategies stopping late in life, i.e., over age 80, yield negative effectiveness (net harm), meaning these strategies are clearly dominated. Screening too frequently, i.e., more often than triennial screening, do not appear on the efficient frontier in the base-case of the example presented.

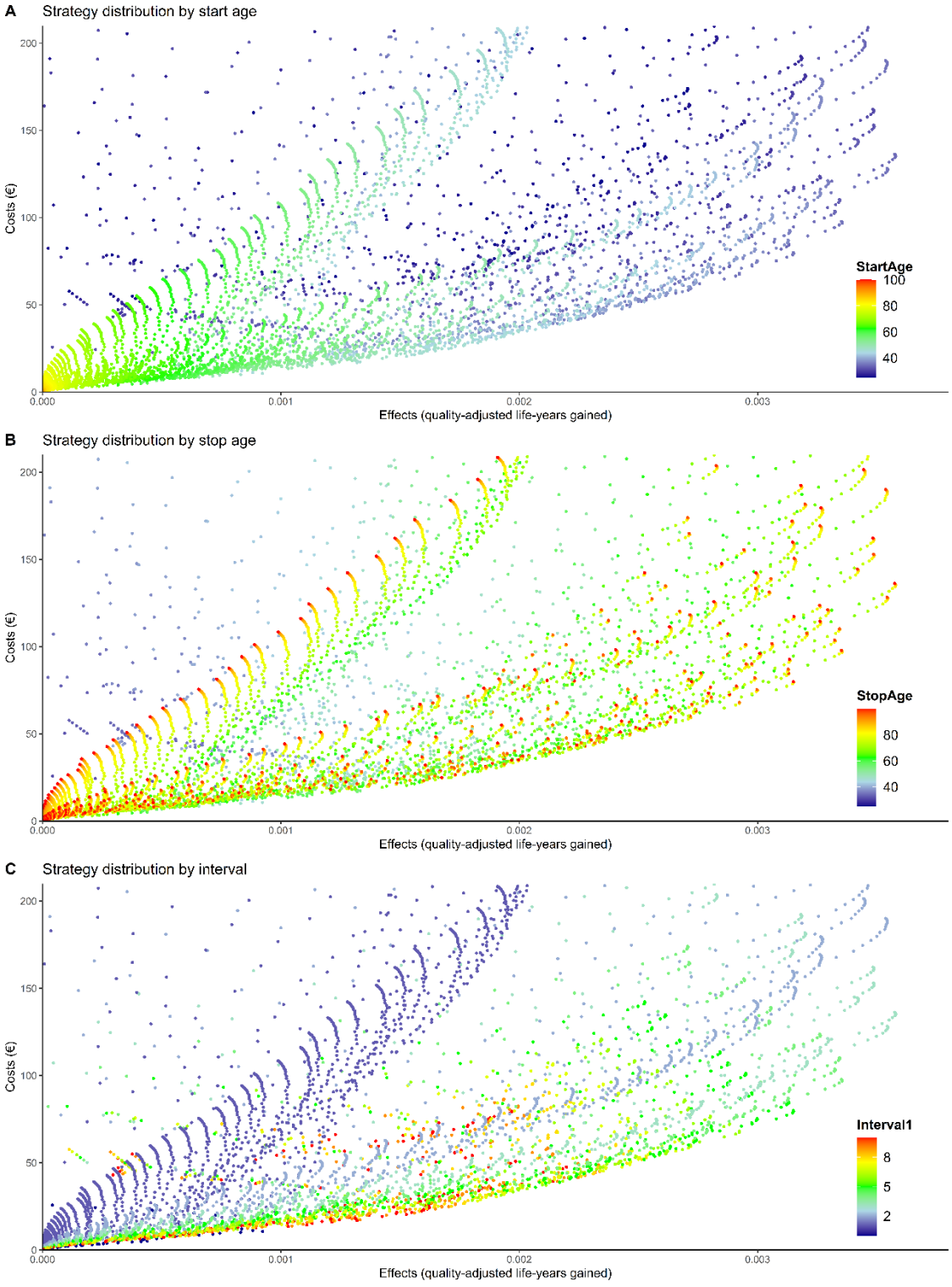
To more easily observe the composition of the efficient frontier this study also includes Figure 4-2 with a more appropriate axis scale. The range of strategies on the efficient frontier in the simulated model varies from strategies with low ICERs, narrow age ranges and long screening intervals up to those with higher ICERs, broader wider age ranges and shorter screening intervals. The detailed characteristics of strategies on the efficient frontier can be found in the second study of this thesis (Table 3-1). Naturally these results are contingent on the particular input parameters used in this example and will likely differ with different assumptions about the natural history of disease, test performance and other model parameters.

Figure 4-1 Screening strategies with common characteristics on the cost-effectiveness planes



This diagram plots the cost-effectiveness results (same estimates as in Figure 3-10), using different colours to illustrate the distribution of strategies with distinct values in screening starting age (Panel A), stop age (Panel B), and interval (Panel C). It features 8,161 plotted strategies, including the no-screening scenario, for the base-case analysis. The strategies cover screening intervals from 1 to 10 years, with start and stop ages ranging from 25 to 100. The cost and effectiveness estimates are derived from the model proposed in the second study of this thesis.

Figure 4-2 Screening strategies with common characteristics on the cost-effectiveness planes (relative to no-screening, partial results rescaled for clarity)



This diagram presents results identical to those in Figure 4-1, but it incorporates adjusted x- and y-axes. The cost and effectiveness estimates are referenced to the no-screening option, positioning the no-screening point at the origin of the cost-effectiveness plane. The purpose is to highlight patterns, when strategies share the same value for one aspect (starting age, stop age, or interval) while varying the other two aspects, particularly for the range of strategies closer to the efficient frontier—something that is scarcely observed in Figure 4-1. Each dot on the diagram represents the cost and effectiveness estimates of every screening strategy in the base-case analysis.

The full analyses in Figure 4-1 and Figure 4-2 show the range of strategies as all of the strategy characteristics of screening start age, stop age and the interval are varied. It is difficult to observe how costs and effects vary as all of these characteristics are changed simultaneously. Accordingly, this study also plotted the cost-effectiveness planes from a subset of strategies in Figure 4-3, holding two out of three characteristics of screening schedules constant while varying the third. This permits a more transparent depiction of costs and effects vary with changes to screening characteristics. The following paragraphs summarised how the positions of the strategies vary in the cost-effectiveness plane characteristics of screening are adjusted using the illustrative example shown in Figure 4-3.

Different stop ages

Figure 4-3A demonstrates the variation of costs and health effects when varying the screening stop age while holding the screening start age and screening interval constant. Figure 4-3A has three lines corresponding to three different screening intervals of annual, biennial and screening every ten years. Generally, as the screening stop age increases the costs and effects of screening increase. The increase in costs and effects is initially almost linear as the stop age is increased but eventually costs rise more than proportionately to effects at older stop ages and ultimately increasing the stopping age may result in net incremental harms. In our example (Figure 4-3A), the screening strategies stopping after age 80 offered higher costs but lower effectiveness because the stop ages became too high and the benefit of the additional screening at such a high age was less than the harm it brings. This can be observed by all the lines in Figure 4-3A, especially obvious in the line presenting the annual screening.

The marginal effectiveness and cost are diminished when expanding the age stop age. As shown in Figure 4-3A, screening strategies starting at age 30 had a larger increase in costs or effects if the screening strategies with stop age changed from 40 to 50 years old compared to screening stopped changed from 50 to 60 years old. This increase got smaller when the stop age was higher.

Different start ages

Figure 4-3B illustrates the variation of costs and health effects when varying the starting age while holding the screening stop age and screening interval constant. Figure 4-3B

has four lines corresponding to three different screening intervals, including every 1, 2, 10 years. These lines were only close to linear for middle starting ages, i.e., between age 40 and 80. The strategy with a lower screening starting age (larger age range) has a higher cost estimate and tends to have higher effectiveness if the screening starts not too early and is not too intense. This also depends on if the benefits can outweigh the harm that extra screening brings. In the example (Figure 4-3B), the screening strategy did not gain any effectiveness when it started screening earlier than age 40 if screening annually or every 10 years. The screening commencing later than age 80 has a low increase in cost and effectiveness, which is close to the no-screening scenario.

For strategies starting after age 40, the marginal effectiveness and cost are diminished when the starting age is closer to the stop age. Figure 4-3B shows the screening strategies which stopped at age 90 generally got higher returns in cost and effects when started screening earlier. The screening had the largest gain in costs and effects when screening starting at age 40 compared to any starting ages. Compared to other starting ages, screening earlier than age 40 have higher marginal cost increases but smaller marginal effect increases which are even negative sometimes, i.e., screening annually and every 10 years, making the overall effectiveness lower than screening starting at age 40.

Different screening intervals

Both Figure 4-3C and Figure 4-3D present the variation of costs and health effects when varying the interval while holding the screening starting age and stop age constant. Figure 4-3C has four lines separately for screening with different stop ages at 40, 50, 60, and 90. Figure 4-3D has four lines for screening with different starting ages at 30, 40, 50, and 60. Both figures have the shape of curves close to semi-oval which size is larger as the screening age range increases. These curves have an annual screening on the top left and the screening every 10 years on the bottom left, which were the least and the largest screening intervals in our example (Figures 3-9C and 3-9D). The highest effectiveness happened when screening in a moderate interval, between every 2 and 5 years. Both figures only plot the strategies that the range of screening ages can be evenly divided by screening interval.

Comparing the strategies with the same starting age and stop age, the marginal increase in cost is consistently higher when the screening interval decreases (more intense) where

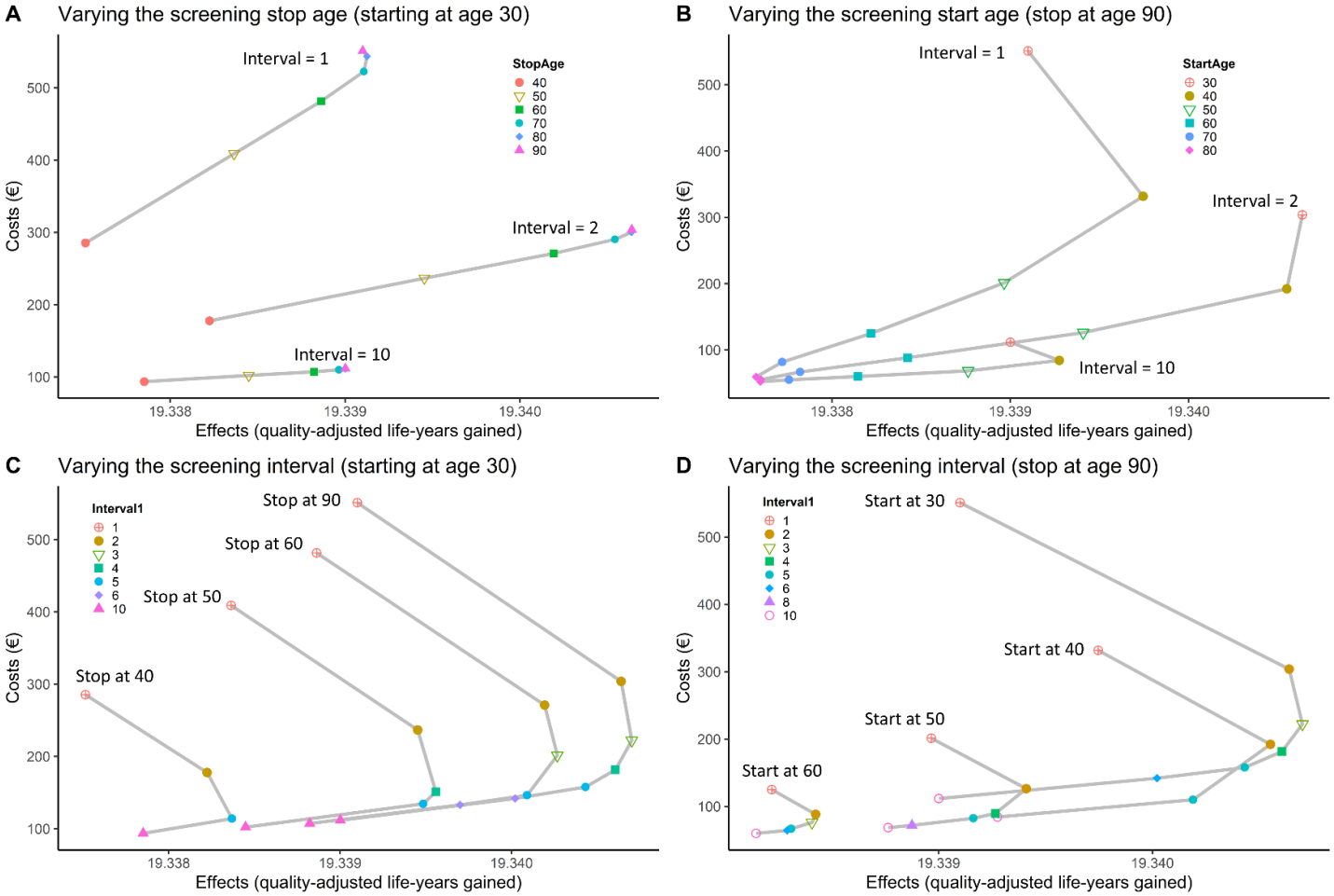
the change in effectiveness is not necessarily stably growing. The marginal effectiveness initially increases when the screening is more frequent until a certain screening interval, typically 2-5 years, and decreases afterwards. When additional screening starts bringing negative marginal effects, the strategy becomes strongly dominated.

Possible strategy omission

If the optimal strategy is the strategy at either the upper or lower extreme of the efficient frontier this may indicate a possible case of relevant strategy omission. If the optimal strategy has the highest effectiveness in the analysis (i.e., is at the upper extreme of the frontier), then the CEA may have omitted most intense strategies, such as those with shorter intervals, higher stop age or lower start ages. In that case, further screening strategies can be added to the analysis by broadening the screening age range or shortening the screening interval. Conversely, if the optimal strategy is the strategy with the lowest cost (i.e., is at the lower extreme of the frontier), then the CEA may have omitted lower intensity strategies, such as those with longer screening intervals and narrower screening age ranges.

This study suggests more strategies should be simulated until the screening is frequent enough and a negative return is observed from additional screening service. For example, an annual strategy starting at age 30 reaches its highest effectiveness when the screening stops at age 80. This means any annual screening strategies end before 80 are worth considering. This rule can be applied to constant screening stop age and interval.

Figure 4-3 The influence of screening schedules on the cost and effectiveness of screening programmes



These are the partial results of the base-case analysis. Diagram A illustrates strategies that start at age 30 but have varying stop ages. Strategies with the same screening interval are connected to observe the influence of stop age under different screening intervals. Diagram B shows strategies that stop at age 90 but start at different ages. Strategies with the same screening interval are linked to observe the influence of start age under different screening intervals. Diagrams C and D display strategies starting at age 30 and stopping at age 90, with connected strategies having the same screening stop or start age, respectively.

The efficient frontier

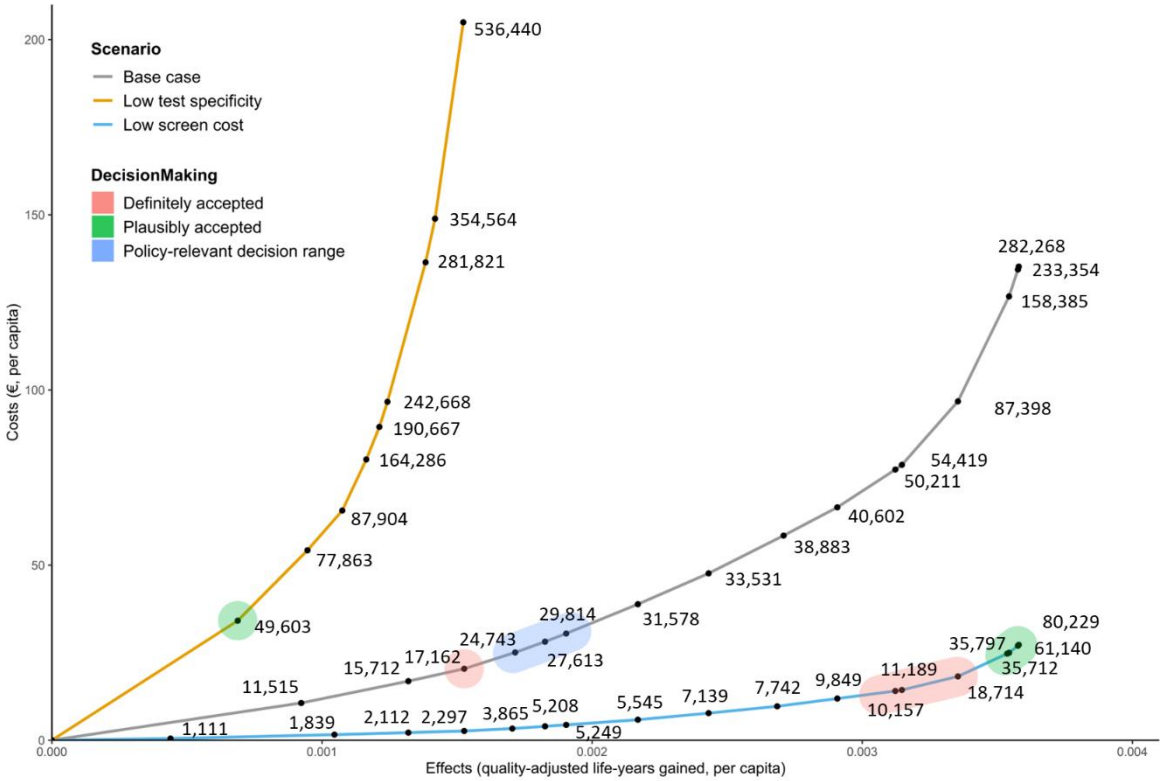
Smooth frontier

The ICERs on the efficient frontier are a series of slopes between two screening strategies in increasing order. The first ICER has the lowest slope at the beginning of the efficient frontier and the slopes get higher as the effectiveness is higher. The potentially very large number of strategies that can apply in screening means CEAs can yield efficient frontiers with smoothly curved shapes rather than a small number of discrete increments between a small of strategies. In principle, increasing the number of strategies considered should reduce the difference between successive increments on the efficient frontier and the increase in ICERs along the frontier should be relatively gradual. Simulations that include very few of the potentially available strategies may result in frontiers composed of few strategies and feature visibly noticeable kinks as ICERs increase rapidly between successive strategies.

Policy-relevant portion

Ideally, the efficient frontier should cover the range of decision-making threshold. This policy-relevant portion of the efficient frontier is the main part of the research results that can influence the decisions. Figure 4-4 shows the location of the policy-relevant portion of the efficient frontier and its implication for decision-making. The base-case scenario has three screening strategies located within the policy-relevant portion of the efficient frontier and the strategies just above and below the thresholds are not too far away from the threshold. This example provides a good selection of strategies for decision-making. In the scenario of low test specificity, the frontier is steeper and its lowest ICER is €49,603, yielding a large gap between the optimal strategy and threshold. In the scenario of low primary screen cost, a large portion of the efficient frontier consists mainly of ICERs below the range of the decision-making threshold and it does not capture the decision-making threshold well. CEAs should explore more strategies to fill the gap and to feature more strategies within the policy-relevant portion of the efficient frontier, thereby ensuring a smoother curve on this portion of the frontier.

Figure 4-4 Policy-relevant decision range of the efficient frontier



The plotted efficient frontiers are for illustrative purposes and do not correspond to any real-case analysis; they are derived from the pedagogical model presented in the second study of this thesis. Ideally, the frontier should encompass the range of threshold values and the policy-relevant decision range. Strategies with incremental cost-effectiveness ratios (ICERs) below the threshold may not necessarily maximize health benefits, and strategies with ICERs above the threshold should not be accepted under the general decision rule. However, in cases of low test specificity and low screening costs, where no strategy falls within the policy-relevant decision range, an ICER slightly higher than the threshold may be considered as a possible best policy.

Moreover, the chance of selecting any given strategy as the optimal strategy is higher when larger difference in ICERs observed between the strategy and the subsequent strategy on the efficient frontier. This is particularly likely to occur when there is no explicitly stated cost-effectiveness threshold or the threshold is not robust and might be negotiable. In the scenario of low test specificity, the strategy with an ICER of €49,603 is chosen under a wide range of thresholds. The threshold can be ranged from €49,603 to €77,862, which is below the ICER of the subsequent strategy on the efficient frontier. Conversely, in the base-case scenario, the strategy with an ICER of €29,814 is only optimal when the threshold ranges between €29,814 and €31,577. Because the ICERs of adjacent strategies are more likely to be similar when many strategies are simulated the estimated differences in costs and effects will be small if one is chosen instead of another. Conversely, if few strategies are simulated the estimated differences in outcomes may be quite large if one is chosen in preference to another. This also specifies the importance

of identifying strategies on the efficient frontier slightly beyond the range of provisionally explicit thresholds which is beneficial for decision-making in practice.

Frontier convexity

The optimal strategy is normally the point on the efficient frontier that is tangential to the slope of the threshold. Accordingly, the convexity of the policy-relevant portion of the efficient frontier is influential on the optimal strategy. The degree of convexity of the frontier depends on whether cost and health effects vary proportionately as screening intensity increases. This degree of proportionality can depend on parameters in the model, resulting in more or less convexity with different parameter values. When the policy-relevant portion of the efficient frontier is closer to linear (not convex), an omission of any given strategy is likely to have less influence on ICERs than when the frontier is more convex. Further examples of this policy significance are considered below.

Parameter impact

This study observed the influence of changes in the preclinical duration and test sensitivity on the degree of convexity of the efficient frontier (Figure 4-5 and Figure 4-6). These parameters make some strategies more cost-effective and others less cost-effective, resulting in inconsistent impacts on the intensity of the optimal screening strategy over a range of different threshold levels.

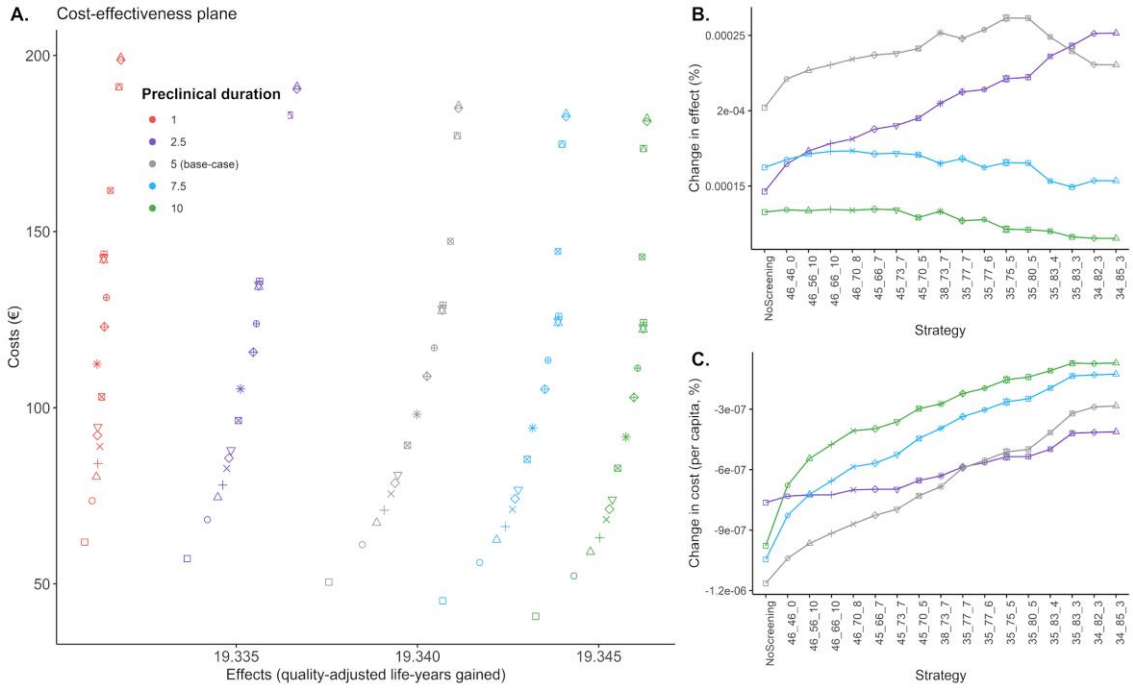
Preclinical duration

A longer preclinical duration results in low-intensity screening strategies having lower ICERs and high-intensity screening strategies having higher ICERs. When the preclinical duration is short, additional screening yields QALYs gain and cost savings. Conversely, with a long preclinical duration, the marginal effectiveness decreases and the marginal cost increases by adding more screening. The strategies with moderate intensity have the largest increase in effects. Accordingly, the strategies in the middle of the efficient frontier become more cost-effective as the preclinical duration gets longer.

Test sensitivity

An improvement in test sensitivity makes the efficient frontier more convex, favouring moderate-intensity screening. When the test sensitivity is low, additional screening increases the effectiveness no matter how frequent screening is. The overall costs of screening strategies also decrease as more screens are conducted, although this study observed a slight marginal increase in costs when the screening is very intense.

Figure 4-5 The inconsistent marginal changes in cost and effectiveness when lengthening the preclinical duration

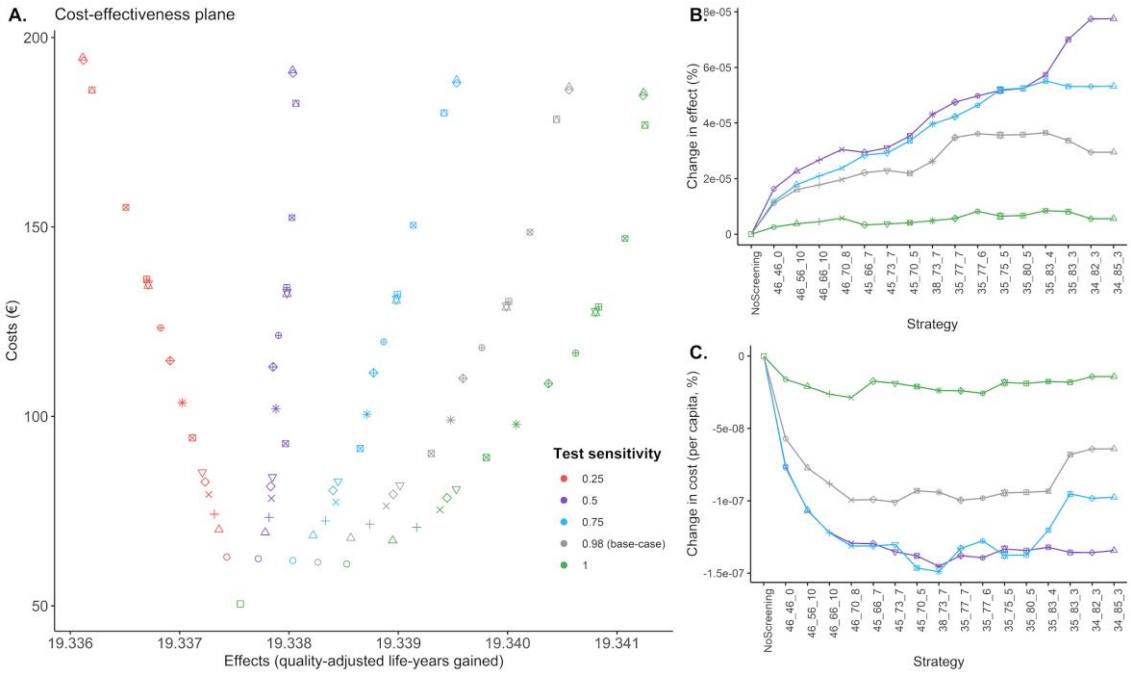


The first number of the strategy's name is the screening starting age and the second number is the stop age. The last number is the screening interval. In Panel A, we plotted the cost and effectiveness estimates for 17 strategies on the efficient frontier in the base-case analysis, where the mean preclinical duration is assumed to be 5 years. Panel B and Panel C respectively illustrate the percentage change in effectiveness and cost estimates across different scenarios. The purple dots/line represents the estimates from the scenario with a mean preclinical duration of 2.5 years compared to the scenario with a duration of 1 year. The grey dots/line represent the base-case analysis compared to the scenario with a preclinical duration of 2.5 years, and so on. To calculate the percentage change, use the formula: Percentage Change = (Difference ÷ Original Number) × 100. The change is determined by finding the difference between the chosen scenario and the next scenario with a longer preclinical duration, divided by the chosen scenario. The change in the convexity of the efficient frontiers happens when the cost and effectiveness estimates do not change consistently across the strategies ordered by the screening intensity. In this example, the grey line in panel B is not linear, showing screening strategies with "mid-intensity" have a higher increase in effectiveness when lengthening the preclinical duration.

When test sensitivity is high the marginal gains of adding more screening are increasingly small as screening intensity increases. Once the screening is relatively intensive further increases in intensity yield very little additional benefits as most cases are detected at early screening rounds and there is little disease left to detect as additional screens are added to a programme. This study observed a lower marginal increase in effectiveness among high-intensity screening strategies. This might be an outcome of the trade-off

between the LY gained from more cases detected earlier and the QALY loss from excessive screening and overdiagnosis. Similarly, screening can save more costs if introducing extra screening when its intensity is low. When the screening intensity is high, then adding addition screening intensity results in more that proportionate increases in costs. Overall, the change in the convexity of efficient frontiers is observed when there is an improvement in test sensitivity.

Figure 4-6 The inconsistent marginal changes in cost and effectiveness when improving the test sensitivity



The first number of the strategy's name is the screening starting age and the second number is the stop age. The last number is the screening interval. In Panel A, we plotted the cost and effectiveness estimates for 17 strategies on the efficient frontier in the base-case analysis, where the test sensitivity is assumed to be 0.98. Panel B and Panel C respectively illustrate the percentage change in effectiveness and cost estimates across different scenarios. The purple dots/line represent the estimates from the scenario with a test sensitivity of 0.5 compared to the scenario with 0.25. The blue dots/line represent the scenario with a test sensitivity of 0.75 compared to the scenario with a test sensitivity of 0.5, and so on. To calculate the percentage change, use the formula: Percentage Change = (Difference ÷ Original Number) × 100. The change is calculated by taking the difference between the chosen scenario and the next scenario with a higher test sensitivity, divided by the chosen scenario. The change in the convexity of the efficient frontiers happens when the cost and effectiveness estimates do not change consistently across the strategies ordered by the screening intensity. In this example, the grey line in panel B is not linear, showing that screening strategies with "mid-intensity" have a higher increase in effectiveness with an improvement in test sensitivity. Both the grey and blue lines in panel C show that "mid-intensity" strategies have higher cost-saving when test sensitivity improves.

Impact of strategy omission

A strategy omission is probably more problematic when the efficient frontier is more convex. The omission of a strategy from a convex frontier can result in large changes in the ICER of the strategy for which the omitted strategy served as a comparator in a more complete analysis. This will be the case if the omission occurs at the portion of the frontier where the ICERs change rapidly between adjacent strategies (i.e., the most

acutely curved portion of the frontier). The consequences of such omission will be less pronounced in examples of less convex frontiers or on more linear portions of the efficient frontier.

Table 4-1 presents an example of the consequences of strategy omission under the scenarios of different preclinical durations. In the scenario of a longer preclinical duration with a more convex efficient frontier, omitting the strategy of screening every 6 years between ages 37 and 73 results in an ICER of €58,745/QALY for the strategy of screening every 6 years between age 37-79. In the case of a complete analysis without the strategy omission, that strategy would have an ICER of €70,950/QALY, which is substantially higher. Conversely, when the preclinical duration is short, the omission of the same strategy results in an ICER of €52,351/QALY screening every 5 years between ages 34 and 74, whereas this is only marginally different at €52,402/QALY in the complete analysis. Nonetheless, a biased ICER estimate is only matters in decision-making terms if the omission has consequences for the policy-relevant portion of the efficient frontier. In both scenarios shown in the examples, the ICERs of optimal strategy are not much different if the threshold is below €40,000/QALY.

Table 4-1 The impact of frontier convexity on incremental cost-effectiveness ratios if missing the relevant strategy screening between ages 37-73 every 6 years

Strategy	Short preclinical duration Average 3 years				Long preclinical duration Average 9 years			
	QALYs	Costs (€)	ICER, €/QALY	Biased ICER, €/QALY	QALYs	Costs (€)	ICER, €/QALY	Biased ICER, €/QALY
41_73_8	139.74342	3,590,248	SD	SD	236.41607	3,760,215	30,030	30,030
44_69_5	185.82516	4,131,239	34,481	34,481	242.19958	4,359,058	ED	ED
37_72_7	208.75217	5,027,272	39,082	39,082	275.41364	5,283,474	39,060	39,060
37_73_6	223.75131	5,740,247	47,534	Missing	288.60168	6,028,838	56,518	Missing
37_79_6	226.31574	5,876,837	ED	48,371	291.00839	6,199,594	70,950	58,745
34_74_5	269.34001	8,129,207	52,402	52,351	301.60863	8,508,788	SD	SD
37_82_5	244.37323	6,954,005	ED	ED	299.70240	7,330,637	130,095	130,095

ED, extended dominated; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year; SD, simple dominated. The first number of the strategy's name is the screening starting age and the second number is the stop age. The last number is the screening interval. In the scenario with a less convex efficient frontier (short preclinical duration), the omission of a strategy results in a similar ICER estimate. Conversely, in the scenario with a convex efficient frontier (long preclinical duration), the difference in ICER estimates is more pronounced.

An example of the proposed method

This study provided an example of an iterative method to expand the selection of strategies with a higher chance to be cost-effective. The details of the five steps are described as follows:

First, for illustrative purposes, this thesis narrowed the scope of the study to only include the screening strategies with a fixed interval length throughout the programme and assumed only one screening modality is available for the disease in question. Therefore, the analysis does not consider strategies changes to the interval length or testing modality over an individual's participation in the screening programme. Furthermore, this study does not consider variation in triage algorithms. More assumptions can be found in the method chapter of this thesis. Strategies for starting screening later than the age of 30 are considered. The maximum screening age is 100 years old. The screening interval is not limited. The thesis only focuses on the screening policy for the general population and does not discuss the strategies for specific risk groups. A decision threshold range of between €20,000 and €30,000 per QALY is assumed in this example.

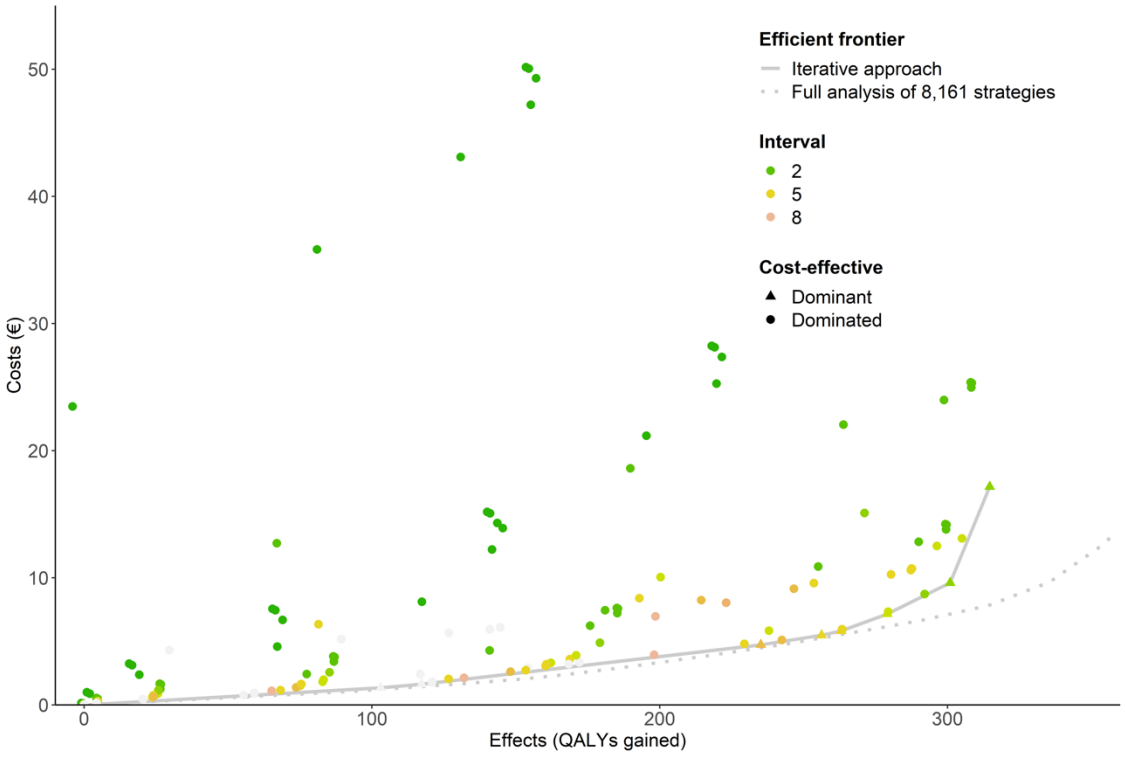
Second, this example chooses the strategies with variations in screening starting age, stop age, and screening interval. This example expects to simulate less than 150 strategies in the first attempt due to the long simulation time. In order to catch a large range of screening intervals but lower the number of strategies, this example equally divided the range into every 10 years. The studied starting and stop ages include 30, 40, 50, 60, 70, 80, 90, and 100, despite other ages available for screening. This age range is chosen because it includes the minimum and the maximum screening age defined in the scope of this example (step 2). Regarding the screening interval, this example chooses every 1-10 years. After removing the screening age range which cannot be evenly divided by the interval, a total of 144 screening strategies are chosen, including the no-screening strategies.

Table 4-2 Cost-effectiveness results of the first attempt

Strategy			Effectiveness (QALYs)	Costs (€)	Incremental effectiveness (QALYs)	Incremental costs (€)	ICER (€/QALY)
No-screening			1,933,756	5,049,812			
50	60	10	1,933,859	6,398,247	103	1,348,435	13,085
50	70	10	1,933,873	6,683,975	14	285,728	20,135
40	70	6	1,933,991	9,782,045	118	3,098,070	26,283
40	70	5	1,934,012	10,532,486	21	750,441	35,581
40	80	5	1,934,019	10,873,775	7	341,289	49,755
40	80	4	1,934,035	12,208,450	16	1,334,675	83,531
40	100	3	1,934,056	14,652,766	21	2,444,316	111,956
30	90	3	1,934,070	22,210,085	14	7,557,319	549,776

ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year. This table shows the results for the strategies on the efficient frontier in the base-case analysis, based on the first attempt of the iterative method. The estimates for effectiveness and costs are discounted. The results of all other strategies can be generated using the model provided in this thesis.

Figure 4-7 The cost-effectiveness plane of the first attempt

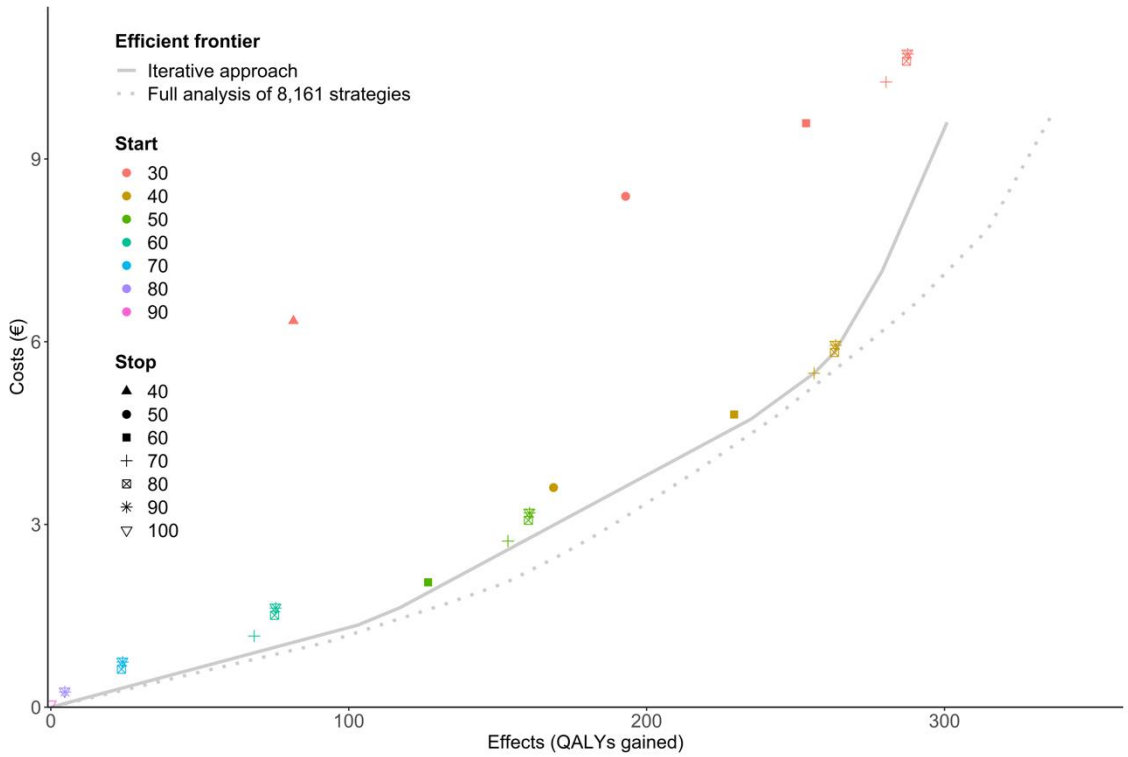


QALY, quality-adjusted life-year. This diagram plots the results of the first attempt, referencing back to the no-screening scenario. A visible gap exists between the efficient frontiers of the iterative approach and the full analysis. The strategies with screening intervals of five years or longer are closer to the efficient frontiers.

Third, the cost-effectiveness results of the selected strategies are observed on the cost-effectiveness plane (Figure 4-7). This study discusses three variables associated with the variation of screening strategies separately, namely screening interval, start age, and stop age. Regarding the screening interval (Figure 4-7), strategies with shorter intervals, e.g., biennial screening, have higher costs, generally distributed on the upper portion of the

cost-effectiveness plane. Considering the decision threshold, the screening strategies on the observed efficient frontier that have ICERs within the policy-relevant range have intervals longer than five years (Table 4-2). Figure 4-8 presents the selected strategies with a screening interval of five years on the cost-effectiveness plane. Regarding the stop age, the screening that stopped later than age 80 has a very small increase in effectiveness, despite a large increase in costs. Strategies starting aged between 40 and 60 cover the range of decision thresholds. As a result, this step identifies additional 53 screening strategies with a higher chance of being cost-effective: starting between ages 40-60, stopping earlier than 80, and having an interval of 5-10 years.

Figure 4-8 The cost-effectiveness plane of the first attempt (partial results: rescaled and some strategy markers omitted for clarity)



QALY, quality-adjusted life-year. This diagram plots the selected strategies with a screening interval of five years on the cost-effectiveness plane. A clear visible gap exists between the efficient frontiers of the iterative approach and the full analysis. In terms of the stop age, screening strategies that stopped later than age 80 show only a very small increase in effectiveness, despite a more noticeable increase in costs.

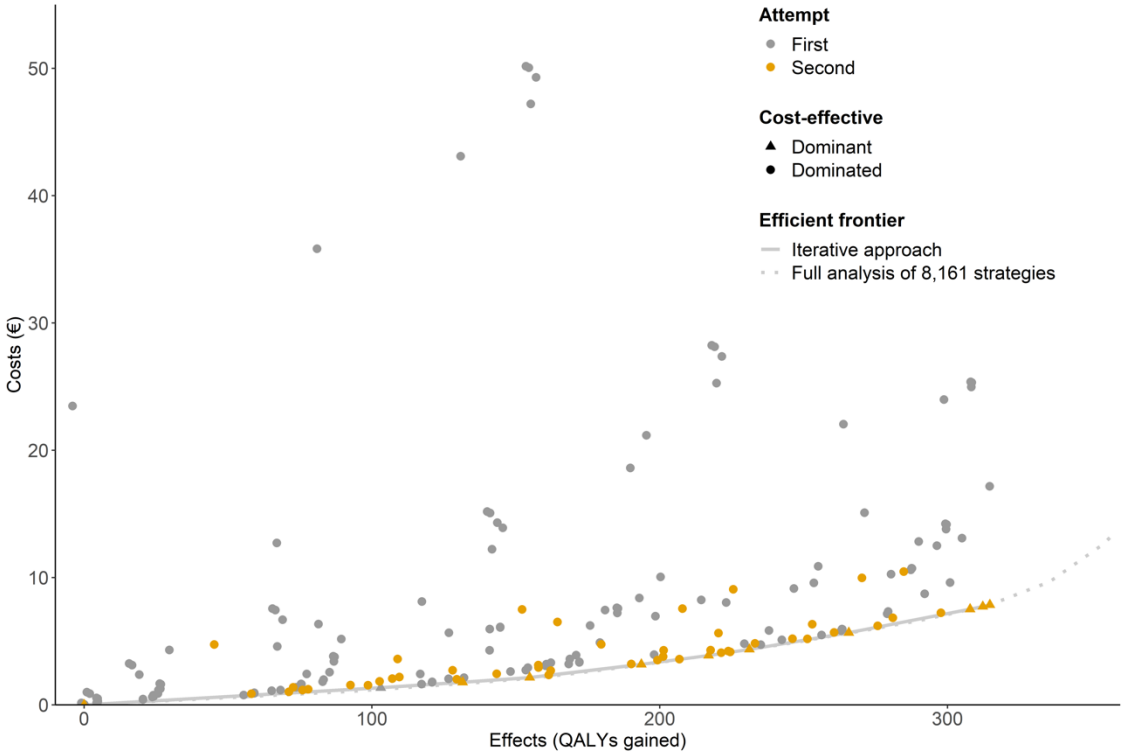
Four, the cost-effectiveness results of strategies identified in the third step are combined with the results of 144 screening strategies in the second step. The efficient frontier is updated with additional strategies in the third step (Table 4-3). The cost-effectiveness plane is plotted with both the first and second attempts of strategy choice (Figure 4-9).

Table 4-3 Cost-effectiveness results of the second attempt

Strategy			Effectiveness (QALYs)	Costs (€)	Incremental effectiveness (QALYs)	Incremental costs (€)	ICER (€/QALY)
No-screening			1,933,756	5,049,812			
50	60	10	1,933,859	6,398,247	103	1,348,435	13,085
45	55	10	1,933,887	6,830,381	28	432,134	15,191
45	65	10	1,933,910	7,201,046	23	370,665	15,907
45	80	7	1,933,949	8,228,864	39	1,027,818	26,512
45	70	5	1,933,972	8,934,439	23	705,575	30,190
35	75	10	1,933,987	9,412,415	15	477,976	33,650
35	70	7	1,934,021	10,718,889	34	1,306,474	37,828
35	70	5	1,934,063	12,569,330	42	1,850,441	43,900
35	75	5	1,934,068	12,776,848	5	207,518	47,149
35	80	5	1,934,070	12,910,619	2	133,771	54,419

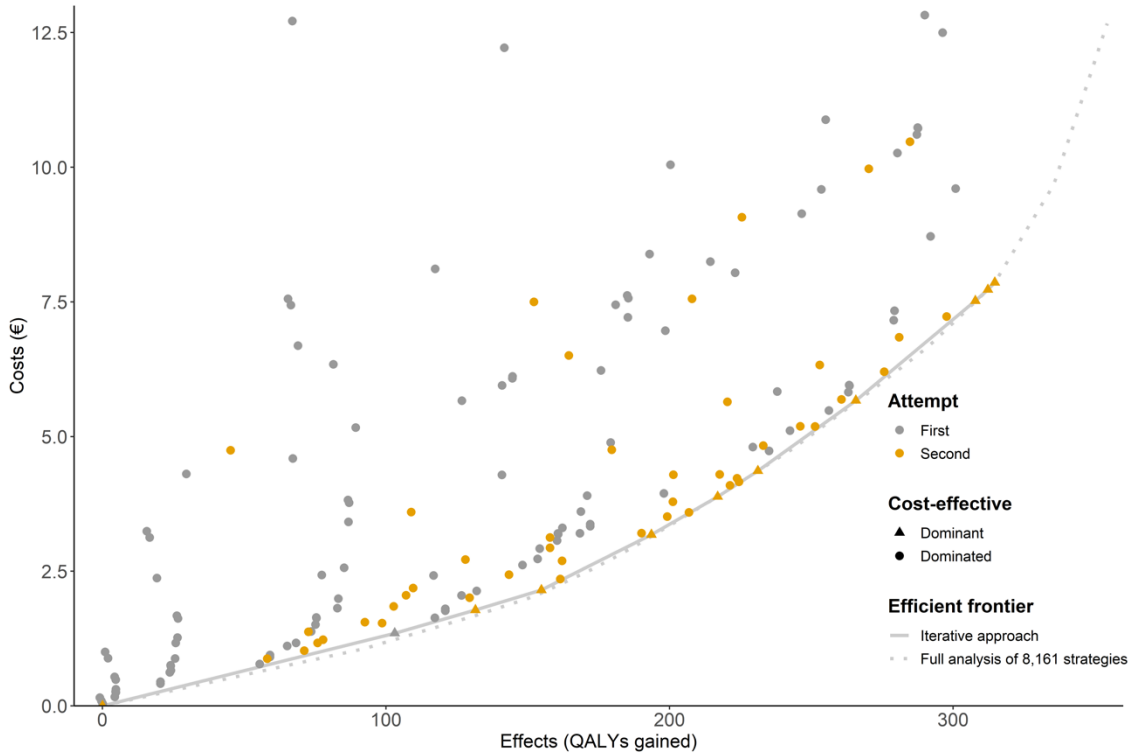
ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life-year. This table only shows the results for the strategies on the efficient frontier in the base-case analysis, combining the results of the first and second attempts using the iterative method. The estimates for effectiveness and costs are discounted. The results of all other strategies can be generated using the model provided in this thesis.

Figure 4-9 The cost-effectiveness plane of the first and second attempt



QALY, quality-adjusted life-year. This diagram plots the results of the first and second attempts, referencing back to the no-screening scenario. The results of the second attempt are closer to the efficient frontier and successfully fill in the gap between frontiers. After strategy expansion, no visible gap exists for most parts of the efficient frontiers between the iterative approach and the full analysis.

Figure 4-10 The efficient frontier of the first and second attempt



QALY, quality-adjusted life-year. This diagram plots partial results of the first and second attempts, with a narrower axis range to better understand the distribution of the strategies closer to the efficient frontier. The results of the second attempt significantly approach the efficient frontier and effectively fill in the gap between frontiers. After strategy expansion, no visible gap exists in the middle part of the efficient frontiers between the iterative approach and the full analysis.

Fifth, stop simulating extra strategies as the policy-relevant portion of the efficient frontier is a smooth curve which ICERs increase gradually. Figure 4-10 shows that our second attempt is much closer to the "real" efficient frontier, and successfully fills some of the gaps in the efficient frontier of the first attempt. Although the second attempt is not exactly the same as the results of simulating 8,161 strategies, this iterative method provides an efficient way of approaching the ideal ICERs with a small number of strategies.

Discussion

Brief findings

This study addressed three principal factors relating to the identification of optimal strategies in CEAs of screening. First, this analysis examined the distribution of screening strategies within the cost-effectiveness plane according to the strategy characteristics. Second, this study introduced the concept of the policy-relevant portion of the efficient

frontier and explained the importance of ensuring the scope of strategies covers it. Third, this study illustrated the convexity of efficient frontiers and examined if it varied with parameter changes. In our model, only changes to test sensitivity and the preclinical duration altered the convexity of the efficient frontier. This study provided some examples of possible screening strategies omission and a practical guide on how to best expand the selection of strategies in a screening CEA using an iterative process.

This study demonstrates the relevance of complete strategy comparisons through the application of a pedagogical model detailed in the second study of this thesis. An advantage of applying the simple pedagogical model is that all the integer variations (i.e., with screening intervals and start and stop ages of whole numbers) of all possible screening strategies (within certain constraints) can be analysed, which is typically very demanding in applied CEAs. From the observation of thousands of strategies, this study made generalised findings and identified situations that likely involve the omission of the optimal strategy and/or relevant comparator strategies. The pedagogical model provides us insights with which to understand how the cost-effectiveness estimates might vary as factors within the model vary.

Compared to the first study of the pedagogical model (the second study of this thesis), this study provides additional guidance on how to avoid the omission of relevant strategies and interpret the impact of parameters on convexity of the efficient frontier. The second study of this thesis provided details of a pedagogical model and illustrated the position of the efficient frontiers on the cost-effectiveness plane when the parameter value changes. This third study extended the observation of efficient frontiers to their convexity and discussed the change in ICERs on the efficient frontiers as input parameters change. This study further explained why the policy-relevant portion of the efficient frontier is important for the identification of optimal strategy.

How to achieve a smooth frontier

This study revealed the association between the characteristics of screening strategies and their cost-effectiveness estimates. In our model, holding the screening start age and screening interval constant, increasing the screening stop age results in an almost linear expansion of cost and health effects up until more extreme stop ages. Similarly, holding

the stop age and interval constant but reducing the start age results in an almost linear increase in costs and effects until the lowest starting ages. Holding the screening age range constant and varying the screening interval yields a skewed bell-shaped curve. Similar patterns are observed in various scenarios in OWSA. In addition, the study disclosed the inconsistent marginal changes in costs and effectiveness estimates of strategies when lengthening the screening age range or increasing screening intensity.

A possible strategy omission can be inferred from the observation of the CEA results. If the CEA does not have a smooth efficient frontier in which ICERs cover the range of the decision threshold, this CEA is likely to have under-explored the range of strategies. The observed repeated patterns on the cost-effectiveness plane indicated the projected trajectories of the strategies and the characteristics of the next strategy can be inferred from this trajectory. If the range of ICERs of the efficient frontier is exclusively above the decision threshold, then there may be scope to explore other strategies with lower intensities such as strategies with shorter screening age ranges or longer screening intervals. In contrast, if the range of ICERs is exclusively below the decision threshold, it may be better to explore the strategies with higher intensities (shorter intervals, lower start ages and higher stop ages). The distribution of strategies on the cost-effectiveness plane gives the analysts clues to expand their selection of strategies.

This study did not find literature explicating the concept of the policy-relevant portion of the efficient frontier and the need to inspect the simulation results. The first study of this thesis, a systematic review, only found 23% CEAs of CRC screening having ICERs that ranged from below to above the relevant cost-effectiveness threshold. Many CEAs do not come close to estimating a full frontier with curves as they only have a small number of strategies on the efficient frontier. Only 3% CEAs identified more than 10 strategies on the efficient frontiers and most of them only have two strategies (35%) or three (n = 20, 25%) strategies. It is obvious that many existing CEAs did not try to estimate a full efficient frontier which is highly likely to have a more cost-effective strategy omitted.

The review in Appendix 1 summarised the screening strategies included in the existing CEAs of CRC screening. This review found only 33% CEAs assessed annual screening and all the CEAs that included annual screening intervals found it to be optimally cost-effective. This might imply those who only simulated biennial screening should consider

the possibility of an annual screening strategy being more cost-effective. The findings of the optimal age range indicate that many CEAs did not go beyond what they found to be optimal. For example, if CEAs find starting at age 50 to be optimal, CEAs should ensure the evaluation of the strategies starting at 45. This review found many CEAs stop simulating more strategies and usually leave the characteristics of the optimal strategy to stay at the edge of the range they define for the screening strategies.

Because of an observed trajectory of strategies with common characteristics on the cost-effectiveness plane, this study is able to propose a more feasible solution to the issue of strategy omission in the CEAs of screening. The proposed method for the avoidance of strategy omission is to observe the association of strategy features and their cost-effectiveness estimates. Expanding the selection of strategies to ensure that strategies with the chance of being on the efficient frontier are considered in the analysis. The consideration also includes the ICER estimates to apprehend which direction of the strategies should expand on the cost-effectiveness plane. For example, analysts may model parts of the efficient frontier where the ICERs are very high. In this case, analysts should consider strategies with less frequent screening.

The proposed method requires a few attempts to include all the possibilities in the scope of the study. Ideally, the CEAs should exhaust all the strategies which have a chance to be cost-effective in their simulation. This, however, is somewhat impractical due to many reasons. There is no clear guide on when to stop exploring the strategies, but this study lists some considerations that might be useful for analysts to decide when to stop.

First, the scope of the study. If analysts clearly know the boundary of chosen strategies, CEAs certainly can narrow their inclusion of strategies. For example, this study only simulates individuals till age 100. The age of 100 is the boundary of the screening stop age.

Second, the restriction of resources in modelling, including labour, time and hardware. The modellers might not have capability to optimise the models or improve the efficiency of the simulations. A long simulation time makes the simulation of a large number of strategies or attempts infeasible especially as most research projects have due dates. Considerations relevant to the hardware include the number of available machines, how

efficient are the machines, and the size of memory. For instance, saving intermediate outcomes might occupy a large space of memories in the device.

Third, other considerations from the perspective of decision-makers. Governments may favour screening strategies which have an integer starting age, stop age, and screening intervals. Resources constraints and the feasibility in the real-world setting are also worth being considered.

In summary, analysts know better when to stop expanding the selection of strategies as it largely depends on the situation and constraints. The most important thing is that the analysts should understand and specify the limitations and consequences of excluding specific strategies from the CEA.

This study did not discuss the screening modality in the proposed method as this study regards the characteristics of modalities as parameters, although the choice of screening modalities is also an important element of screening strategies. From the observation in the second study of this thesis, an improvement in test sensitivity will move the frontier toward the right, making all the screening strategies more cost-effective. If multiple screening modalities with the same or very similar test specificity and the same screening schedules are compared, the screening strategies with higher test sensitivity will be more cost-effective. Accordingly, in the case of a small difference in test specificity, the study suggests simulating the screening modalities with the same combinations of screening starting ages, stop ages, and intervals in the first attempt to identify which is the ideal screening modality. After the identification of the ideal modality, the analysts may determine the ideal screening schedule for a specific screening modality with the proposed method. Future research can develop a more practical guide for the choice of screening modalities in the CEAs as most modalities have variations in both test sensitivity and specificity.

This study provides a definite solution to the research problem of strategy omission in the CEAs of screening in accordance with the suggestions of Torrance et al. (1996). Torrance et al. (1996, pp. 65-66) advised the consideration of variation of intensities and CEAs should use the next-less-intense screening strategy as comparators. An example in Torrance et al. (1996) is that the calculation of ICER for the screening every 3 years

should be compared to the screening strategy every 2 years. Aligned with their suggestion, this study gives more details of other dimensions that were not clearly described in Torrance et al. (1996), including starting ages and stop ages, based on the concept of the next-less-intense screening strategy. CEAs should also consider variations in the screening ages in addition to the screening intervals.

Regarding the boundary of the choice of strategies in the CEAs, Torrance et al. (1996) suggested that analysts adopt the principle of including all the strategies that are really feasible in the CEAs. The proposed method in this study has the same suggestions in Torrance et al. (1996) that analysts have the best judgement to decide when to stop including more strategies. "The analyst will have to decide at what point comparisons are no longer realistic and when programs differ little enough in their effect that finer distinctions will not offer much additional insight (Torrance et al., 1996, p. 66)." For example, analysts can judge whether to adopt screening every half year, if annual screening is included in the analysis.

This study did not find any guidelines or textbooks that have detailed methods or examples illustrating the management of strategies in the CEAs of screening. Although Torrance et al. (1996), compared to others, provided more explicit guidance on the inclusion of screening strategies, they did not provide details on how to conduct a CEA with their theories. In practice, the simulation of all the variations is very demanding for CEAs. The iterative method recommended by this study can overcome this difficulty and align with the suggestions of Torrance et al. (1996).

Decision-making with efficient frontiers

This study found only parameters that directly influence the number of screen-detected cases have an impact on the convexity of the efficient frontier, namely preclinical duration and test sensitivity. When preclinical duration is longer, the screening has more chances to detect the disease before it is symptomatically presented. When test sensitivity is improved, screening has a higher chance to detect disease. Both scenarios favour low-intensity screening. In contrast, a lengthening of preclinical duration has high-intensity screening detect the same number of cases as low-intensity screening but have unnecessary screens. The improvement in test sensitivity has the first few screens in the

high-intensity screening strategy have detected the preclinical cases and the later screens cannot detect cases but bring harms. Notably, the preclinical duration is normally not controllable, but we can choose the screening modalities.

The convexity of the policy-relevant portion of efficient frontiers is critical for decision-making as this influences the robustness of the optimal strategy under different thresholds. When the efficient frontier is more convex, strategies with high intensities have ICERs increase and strategies with low intensities have ICERs decrease. This means the low-intensity screening becomes more cost-effective and high-intensity screening becomes less cost-effective. It is important to identify more strategies when the policy-relevant range of efficient frontiers is more convex as the ICER of the optimal strategy is more different than the strategies with high intensities. This optimal strategy is also more robust in this case, because it is chosen under a larger range of threshold values. Hence, with a higher test sensitivity, analysts may need to be more careful including more screening strategies in their CEAs as the policy-relevant portion of the efficient frontier has a higher chance of being more convex. In contrast, when the policy-relevant portion of the efficient frontier is less convex, the optimal strategy is more replaceable by other strategies which offer similar ICERs.

It is important to find a smooth efficient frontier which has an ICER closer to the threshold which maximises the net benefits (Drummond et al., 2015). The first study of this thesis observed very few CEAs had a smooth efficient frontier as they only included the screening strategies that appear in the clinical guidelines or the existing literature. This means they might ignore optimal strategies and have biased ICER estimates. Their ICERs can merely present the improvement of cost-effectiveness between two interventions which is not the real aim of many CEAs of screening. National screening programmes are often funded by jurisdictions, so the design of CEAs should match their expectations which is to find the real optimal strategy from the societal perspective. This study intends to address the problem of under-exploring the strategies and substandard decision-making with biased ICERs.

Compared to empirical observation

This study can only provide a general guide in choosing strategies in cancer screening CEAs because different diseases have very different disease histories and available screening modalities, resulting in various shapes of the efficient frontier. For example, cervical screening is found to have a more convex efficient frontier (van Rosmalen et al., 2012) and lung cancer screening is observed in some studies to have a less convex, more linear efficient frontier (Ten Haaf et al., 2017). Also, the tolerance towards the burden of over-diagnosis influences the choice of screening strategy (De Gelder et al., 2011). For example, it is controversial to have a national prostate cancer screening programme due to very high rates of false positives and overdiagnosis because the potential benefits probably do not outweigh the potential harms (Srivastava et al., 2019). In 2018, the United States Preventive Services Task Force recommended leaving the decision of whether to receive screening to personnel after their discussion with clinicians about potential benefits and harms (Grossman et al., 2018). Each cancer CEA has an independent composition for adjusting the frequency of screening to expand the selection of screening strategies.

Similar patterns on the cost-effectiveness plane are also observed from applied cancer screening CEAs. Although many CEAs included an extensive range of screening schedules tend to report partial results only for strategies on the efficient frontier, we can observe their reported ICER still increases as the screening intensities increase with a more extensive age range or shorter screening interval (Goede et al., 2013; Naber et al., 2016; Van Den Akker-Van Marle, 2002). If the same screening schedules are applied, the characteristics of screening modality are the only factor that influences the cost-effectiveness of screening strategies. Increases in test sensitivity has unambiguously favourable outcomes in terms of increased screening effectiveness and reduced net costs, though the impact on ICERs can be ambiguous as the ICERs of some strategies will fall while others may rise. In the CRC screening CEAs applying the fixed screening schedules to all the modalities, the frontier is often composed of strategies with higher sensitivity, e.g., lower cut-off of faecal immunochemical testing (FIT) (Goede et al., 2013; Wilschut, Hol, et al., 2011).

Other methods for identifying optimal strategy

The iterative method offers the advantage of identifying the optimal strategy with significantly fewer simulation runs compared to the brute force method, simulation of all possible strategies. Beyond its efficiency, the iterative approach also holds the potential to enhance the insights of modellers into simulation estimates and contribute to the selection of optimal strategies.

Alternatively, another approach to identify the optimal strategy involves constructing a metamodel, a complex statistical model. However, this method presents challenges due to the intricate nature of defining screening strategies through a series of parameters. The resulting cost-effectiveness outcomes are intricately linked to these parameter configurations. For instance, strategies involving a wide age range and longer intervals probably yield similar cost and effectiveness outcomes as those featuring a narrower age range and shorter intervals.

Implementing a metamodel probably faces challenges, demanding prior data to support its construction. The model's accuracy hinges on the volume of available prior results. This, however, runs counter to the research objective of identifying the optimal strategy within a constrained range. Moreover, metamodels tend to experience substantial increase in computational time with the addition of more parameters. For instance, Koffijberg et al. (2021) indicated that the Gaussian process metamodel entails exponential growth in computation time as the parameter count rises. Additionally, there is currently a lack of consensus on establishing arbitrary criteria to assess whether the complex model achieves adequate accuracy.

Moreover, developing metamodels demands a solid statistical foundation. Typically, only proficient research teams, equipped with skilled statisticians capable of constructing complex models, can adeptly code a screening simulation model. In this scenario, for such research teams, opting for a comprehensive analysis of all potential strategies might prove more efficient than fitting a complex model that demands labour-intensive development and compromises accuracy. Additionally, small research groups may not benefit from this approach due to a lack of a strong statistical background. In contrast, the iterative method remains accessible and well-suited as long as the results can be

visualized on the cost-effectiveness plane. Most health economists or modelers can readily achieve this using familiar software tools such as Excel and R.

Both methods, metamodels and the iterative method, have their own advantages and disadvantages. The iterative method is straightforward to apply and can identify the identical optimal strategy as the brute-force method. However, the iterative method might not be able to explore screening programs that involve varying the screening interval or applied modalities during the course of the screening program, as the intricate relationship between cost-effectiveness estimates and the characteristics of screening strategies. On the other hand, the metamodel potentially offers a solution to this limitation, as it does not rely on subjective judgment to establish the relationship between these factors. Nevertheless, the accuracy of this method remains uncertain, particularly when it is supported by limited data for its development. There is no evidence to confirm whether metamodels can indeed yield the same optimal strategy as the brute-force method. Both the metamodel method and the iterative method are novel in the field and require further research and practical examples to establish their efficacy.

Limitations and uncertainties

The major limitation of this study is that this study did not fully explore potential uncertainties in CEA models which result in the variation of ICER estimates. Usually, stochastic uncertainty, structural uncertainty, parameter uncertainty, and heterogeneity should be discussed when illustrating the CEA results (Briggs et al., 2012).

First, this study did not illustrate the stochastic uncertainty in the cost-effectiveness results. This study chose to use constant random seeds and simulate a large number of individuals in the simulation to lower this uncertainty.

Second, structural uncertainty was not addressed in this study due to the highly abstract nature of the pedagogical model, which limits the feasibility of applying a meaningful reference case. While Briggs et al. (2012) suggest the application of a reference case, the current model lacks a real case-study scenario, making the meaningful application of such a case challenging. Nevertheless, it is worth noting that we observe similar trends in cost-effectiveness results compared to empirical CEAs. This enhances the reliability

of our findings. Recognising the significance of structural uncertainty, NICE (2022) suggests including examples of different ways of categorising health states and various care pathways. To explore structural uncertainty within our simplified model, we conducted a scenario analysis, varying the sojourn time of preclinical and clinical states in the model. We tested extreme values for both health states, and despite the abstraction, our cost-effectiveness results remained validated. Future research can also extend this to calculate the expected cost-effectiveness results with a model average in which each scenario has equal weight, or set up a parameter that assigns different probabilities to scenarios (Bojke et al., 2009). Overall, the generalisability of our findings needs to be tested in a more complex disease history model.

Third, despite this study considering the uncertainty of parameters that affect the cost-effectiveness results in the model, it solely focuses on the parameters included in the simplified model and the isolated effect of changing a single parameter. Other parameters and their combined effects are worth investigating in the future. Specifically, the sensitivity and specificity of certain screening methods are related. For example, the FIT cut-offs for CRC screening can determine the balance between sensitivity and specificity.

This study employed an OWSA to capture the variation in cost-effectiveness estimates resulting from the uncertainty of individual parameter values. Although PSA is not considered in this study, as the main goal is to understand the specific impact of individual parameters, this study provided the cost-effectivenses plane illustrating PSA results in the Appendix 2. PSA combines the effects of multiple parameters, making it difficult to understand how parameter groups relate to ICERs. PSA is typically used to assess the stability of ICER estimates across the entire range of parameters. In our study, the model does not simulate a real-world scenario where parameters are assumed arbitrarily. As a result, conducting a PSA lacks a clear rationale.

Last, the illustration of observed or explained heterogeneity is significant in screening CEAs because the screening strategies usually have commonalities. This study examines if the characteristics of screening strategies have similar cost-effectiveness estimates. This study concentrates on the characteristics that can be defined, including screening starting age, stop age, and screening interval. This study only discussed more basic aspects of strategy selection and did not explore the strategies that vary the screening

interval or applied modalities during the course of the screening programme. This study did not include the consideration of sequential surveillance and the combination of screening modalities.

Based on the results of the base-case analysis, there might be some more strategies on the efficient frontier infrequent than every 10 years, but this study did not explore the possible range of screening intervals over 10 years, e.g., screening every 15 years. Notably, screening every 10 years results in an ICER lying outside the policy-relevant region of the efficient frontier. Consequently, opting for longer screening intervals merely contributes to a more comprehensive frontier, which does not impact screening policy decisions.

Furthermore, the concept of heterogeneity also includes variations among the individuals simulated in the model. However, this study does not explore potential observed patient characteristics, as subgroup analysis is not relevant to the primary goal of this study: identifying the optimal strategy.

In addition to the uncertainty mentioned in Briggs et al. (2012), there is an uncertainty in the decision threshold. With a cost-effectiveness plane, we can easily illustrate the implications of different threshold values on the choice of the optimal strategy. The iterative method proposed in this study investigated the strategies lying around the policy-relevant range of the efficient frontier which is flexible on threshold values and a possible range of thresholds is applicable. Accordingly, the iterative method can be easily customised to threshold analysis on the cost-effectiveness plane. This study only discussed the implications of ICER as this is the main measure of cost-effectiveness outcomes in most CEAs of screening. This study did not consider the implications of other methods for describing the cost-effectiveness outcomes or the impact of parameter uncertainty (PSA) at various threshold values, e.g., NMB and CEAC. Although NMB calculated with different threshold values can be illustrated on the cost-effectiveness plane, which is the Y intercept of the line passing through the optimal strategy with the slope of the threshold value, this study did not illustrate NMB as it increases unnecessary complexity of our illustration on the cost-effectiveness plane.

Future research

The exploration of value of information related to parameter uncertainty holds merit, particularly within the policy-relevant portion of the efficient frontier. The EVPPI for a parameter measures the economic value arising from acquiring precise knowledge about that parameter's uncertainty. This calculation typically entails subtracting the expected value based on the presently available information from the expected value of perfect information for that parameter. In practice, we typically compute the NMB for all alternatives in CEAs. However, employing this method for screening scenarios becomes unfeasible due to the required simulation time. The computation of the EVPPI for a parameter involves nested loops for evaluating the uncertainty of a specific parameter (1000+ iterations) and other parameters (1000+ iterations) to estimate NMB for each strategy, resulting in a computationally intensive process.

This thesis proposes a refined approach to EVPPI calculation, focusing on strategies with a higher likelihood of being optimal, as they significantly influence the expected value of perfect information. Notably, strategies situated on or near the policy-relevant section of the efficient frontier are the ones we consider for decision-making. Another aspect of EVPPI is the choice of interested parameters. If a parameter only changes the end of the efficient frontier without affecting the choice of the optimal strategy, the value of information of such a parameter would be low. Most parameters do not change the convexity and simply move vertically and horizontally with a similar composition of the frontier, also meaning that these parameters might not be worth exploring. Future research could focus on developing methodologies to effectively estimate the value of information for parameters within CEAs of screening.

More discussion is desired to solve the problem of a large number of strategies available in the screening CEAs. This thesis suggests simulating the possible strategies portion by portion but if the simulation is efficient enough, simulating all the possibilities at once might be feasible. A general discussion to differentiate and compare both methods is needed. Overall, we require a new guideline to handle the issues of strategy selection in the screening CEAs.

5. DISCUSSION

Overview

This thesis addresses the issue of strategy choice and comparison in simulation-based CEAs of cancer screening. The methodological concerns surrounding the choice of strategies has been identified in few reviews of the applied literature to date. This thesis characterised and discussed the problem of inadequate strategy inclusion and proposed a possible solution. The first study systematically reviewed CEAs of CRC screening and described the problem of relevant strategy omission in that literature. The second study demonstrated a de novo teaching model of screening. It provided accessible guidance on face validity to illustrate the impact of parameter change on the position of the efficient frontier. The third study applied this model to demonstrate examples of strategy omission that prove policy significance and discussed how to avoid omission by inspecting the distribution of strategies on the cost-effectiveness plane and their estimated ICERs. This third study indicated the importance of not rigidly defining the strategies for consideration at the outset of an analysis but rather adopting an iterative approach that seeks to expand the analysis to include the most relevant strategies based on interim modelling results.

Benefit to decision-makers

This thesis offers several contributions to health economics decision-making. The first study confirms the problem of inadequate strategy inclusion exists in the CRC screening CEA literature and demonstrates why decision-makers should be careful when interpreting existing CEA evidence in that context. Many CEAs do not offer reliable evidence as they do not consider a sufficiently broad range of strategies and employ a narrow scope of study design. The second study provides a template in R Shiny which can be a user-friendly tool for understanding decision-making with cost-effectiveness estimates. The R Shiny tool offers decision-makers the opportunity to easily comprehend the association among the parameter uncertainty, strategy intensity, and screening cost-effectiveness. The third study proposed a new method to for identifying optimal screening strategies when it is not feasible to simply simulate all of the large number of

screening strategies. This has important practical implications for policy makers seeking valid and reliable evidence.

Critical reflections

This section discussed some critical reflections from the observation of CEAs of cancer screening. This thesis particularly addressed the issues of absent HTA guidelines, inadequate reporting, recommendations for choosing strategies, CEAs by research groups, the meaning of ICERs, and the awareness of comparator issues.

No HTA guidelines for screening

Generally it can be expected that analysts will seek to align their CEAs of screening programmes with prevailing health economic evaluation guidelines. Relying on the guidance of HTA authorities to design their studies, however, leads to a problem when the guidance is unclear or incomplete. The introduction of this thesis reviewed the existing guidance from NICE, CADTH, ZIN, HIQA, and PBAC. There is little or no screening-specific guidance provided in the HTA documents, regarding the choice of comparators.

Although the guidance of Torrance et al. (1996) is clear regarding comparator choice in the screening context, it is not an HTA manual. Few CEAs of screening really referred to this guideline when they chose comparators. Among the CEAs included in the systematic review of CRC screening, there is no CEAs reference to either Torrance et al. (1996) or the comparable chapter by Owens et al. (2016) in the updated version of the Washington Panel textbook. An absence of clear HTA guidance on the choice of screening comparators causes the consistent methodological issue of comparator omission observed in the literature on CEAs of CRC screening. It seems likely that similar problems apply in screening for other cancers.

Issues of adequate reporting

Inadequate reporting of results is identified in CEAs of CRC screening according to the self-developed quality assessment checklist. Among CEAs of CRC screening, this thesis found some CEAs lack numeric reporting of cost-effectiveness results, only providing

their outcomes in the figures or only report ICERs. Some CEAs lack sufficient significant figures or only report the results for some strategies, commonly only reporting the results of strategies on the efficient frontiers. In addition to the items included in the self-developed checklist, this thesis also observed that the CEAs of CRC screening do not have a uniform format of reporting.

The lack of a uniform reporting format and appropriate guidance leads to a waste of time for those reviewing, and potentially leads to problems of insufficient transparency. The review conducted within this thesis required considerable time to simply to understand the incremental comparison in each CEA due to the absence of labelling or the reporting of extra outcomes of incremental effectiveness and costs. The review also spent a long time recording the characteristics of screening comparators because the description of the strategies also does not follow a uniform description, table, or label to report their frequencies and age ranges. This reporting issue is more problematic especially when the strategies are with more complex algorithms and different FIT cut-offs. Without clear guidance, CEAs might mistakenly misreport some important aspects of study design or outcomes. This error will influence future CEAs by having substandard evidence as CEAs are likely to choose the same set of screening strategies as comparators. With deficient reporting, future CEAs might misunderstand the study outcomes of old CEAs and follow their choice of strategies.

General recommendations for comparators

CEAs may also inform their strategy selection by reference to those strategies recommended within clinical guidelines. Experts who advise clinical guidelines and analysts of economic evaluation should look beyond the strategies found to be optimal within the existing literature. The calculation of the ICER is sensitive to the comparator against which the strategy is compared and it is important to ensure that the correct interpretation of the ICER is used when reporting such ratios. Because of heterogeneity between studies, a common difficulty of conducting systematic reviews of CEAs is that ICERs cannot be synthesised and reviews cannot make simple conclusions of what is the optimal strategy (Khalili et al., 2020; Pignone et al., 2002). For this reason, the systematic review of Pignone et al. (2002) reported on the range of CERs v.s. no-screening of common screening strategies across CEAs, instead of attempting to aggregate ICERs. A key novel contribution of this thesis is that it addresses the consequences of applying a

different set of strategies as comparators. This thesis recommends future reviews of screening CEAs scrutinise and explaining the variation of strategy characteristics and their apparent impact on cost-effectiveness estimates."

Regarding strategy choice in CEAs of screening, it is difficult to know what is a sufficient variation in strategies. In addition to the self-developed assessment tool used in the review, the modelling work of this thesis informs us on how to understand strategy variation and how it might be best expanded. A general recommendation for CEAs and critical evaluation of the existing literature is that CEAs should ensure they have variation the features of screening programmes simulated such as the start age, stop age and the primary screening and triage modalities. Consideration of very different strategies might place quite a lot of dependence on the estimates of disease progression. For example, estimates of the effectiveness of screening intervals that are very different from those used within clinical trials will be heavily dependent on the assumptions regarding the preclinical sojourn time. Although some may find reliance on modelling uncomfortable due to the absence of trials or significant deviation from trialled interventions, the investigation of a broad range of strategies is necessary. Even if such modelling may not directly inform policy, it might at least in principle inform further trialling.

For conducting a CEA, this thesis recommends evaluating screening modalities first before considering the intervals and age ranges. Ideally, CEAs firstly simulate all the modalities with a manageable number of screening schedules. The screening schedules should cover the ranges of screening schedules defined in the scope of the study. Further details on how to choose screening schedules efficiently can be found in the discussion of the third study in this thesis. Through observing the results, CEAs should be able to identify the screening modalities likely to be optimal. Because the sensitivity of the screening modality usually has a major impact on the cost-effectiveness, the convexity of the efficient frontier, CEAs can narrow the range of screening modalities into a practicable amount.

A good critical analysis of the existing literature is necessary to ensure we know where the choice of strategies appears sufficient and where there may be gaps. Although we can make inferences around what strategies might be most relevant from the previous literature, we must do so with caution. This thesis suggests identifying all strategies in

the scope of study design and embracing the iterative approach, re-running the model, and looking for more efficient strategies, to make the inclusion of all strategies more feasible.

The thesis is informed by the particular experience in the CRC literature, but the issues are slightly different in other cancers. For example, stratification is important in the literature of lung cancer screening (Røe, 2020; Ten Haaf et al., 2021). In the example of cervical cancer, the questions of strategy choice may be largely about restraining policy makers from adopting too intense screening, especially where the established practice has been for very frequent screening and the following the introduction of human papillomavirus vaccines (Tjalma, 2018).

CEAs by research groups

It is notable that some research groups can account for quite a large proportion of published research outputs. For example, the MISCAN group and other collaborators with the Cancer Intervention and Surveillance Modelling Network (CISNET) are prominent in the screening CEA literature. They have their own strengths and drawbacks in their published CEAs of screening. For example, in the systematic review of CRC, the MISCAN group does well in terms of correct interpretation of ICER and simulating a large number of strategies (Goede et al., 2013; Van Der Meulen et al., 2018; Wilschut, Habbema, et al., 2011; Wilschut, Hol, et al., 2011), while the research group of Ladabaum et al., however, reports with less clarity (Ladabaum et al., 2014; Ladabaum & Mannalithara, 2016; Ladabaum et al., 2004). Despite the recognized strengths, MISCAN analyses have weaknesses in reporting too. They tend to only report the results for strategies on the efficient frontier, not for all strategies and they tend to consider only one modality type (Goede et al., 2013; Wilschut, Habbema, et al., 2011; Wilschut, Hol, et al., 2011).

It seems likely that highly productive research groups perhaps understand good practices for considering variations in screening strategies but that this might not be shared by smaller research groups. Indeed, while the MISCAN group typically uses a large number of strategies in their analyses the first study of this thesis found only 20% of CEAs considered over 10 strategies. This means that while some may understand the best methods, the majority of analyses may still be sub optimal in their choice of strategies.

There is a need for a clear demonstration of the consequences of having few strategies considered and a guide to explain how to avoid the problem of strategy omission. The studies in this thesis aim to provide guidance to those who are not familiar with the application of economic evaluation in the context of screening.

The meaning of ICER terminology

The use of ICER or ACER is debatable in CEAs if the analyses intend to express different meanings. The ICER is calculated for the identification of the most cost-effective intervention in a global sense as it is compared to all the other interventions in the analysis. ACER (CER vs no-screening) is applied to demonstrate the improvement in cost-effectiveness compared to a do-nothing scenario. The distinction of the study purpose between overall optimum and improvement (assessment of any change) is subtle. This might not be well understood by many policy makers or laymen, so it is important for CEAs to specify the reason for choosing ACER or ICER, express its meaning carefully, and discuss the impact on the policy in their reports.

This thesis intends to disclose the mis-use of ACER in the CEAs of screening where ICER should be the correct measure. The systematic review of this thesis is conducted from the perspective that a CEA should attempt to find the overall most cost-effective strategy whereas some studies might be assessing a narrower objective of simply trying to find an improvement in terms of cost-effectiveness. This different intention can radically change whether an analysis is considered adequate or not as the use of a CER that is not an ICER might be legitimate for the second, narrower objective of finding an improvement in cost-effectiveness relative to any given starting strategy. The second study using the pedagogical model is also designed from the perspective of attempting to find the optimally cost-effective strategy overall. For the purpose of maximising the benefits of the costs devoted, this thesis advocates the use of ICERs in the CEAs of screening programmes as opposed to ACERs or CERs.

This thesis raises another concern about using ACER in the screening CEAs where the jurisdiction has already had some kind of population-based screening programmes. Based on the guidelines published by HTA authorities, the ideal comparator is usual care and/or standard of care. If this rule is applied to the CEAs of screening, the best choice of comparator should be the existing screening programme, instead of no-screening. In

this case, CEAs should calculate CER v.s. the status quo, not CER v.s. no-screening (ACER). Here, the intention of CEAs is also assumed that they aim to demonstrate the improvement of the new screening strategy compared to the existing screening intervention in the healthcare system.

This thesis emphasises the importance of inclusion of all the strategies in the CEAs of screening, and it is not critical if CEAs have the choice of ACER, instead of ICER. CEAs can choose other measures, such as NMB, to identify the most optimal strategy in terms of cost-effectiveness and have CER describe the improvement in cost-effectiveness. The most important is that CEAs should recognise their real aims, have a corresponding scope of study, and properly include all the relevant strategies. The observation on efficient frontier and ICERs is only a measure for us to explore the scenarios that might have possible omission of comparators.

Awareness of comparator issues

Both editors of journals and decision-makers may not be fully aware of the methodological concern of strategy omission in the CEAs of screening. This thesis observes many published CEAs have the problem of comparator omission, although they are already peer-reviewed before publication. As this thesis acknowledged, very few reviews pointed out this problem (Fabbro et al., 2022; O'Mahony et al., 2015). Furthermore, among all the checklists and HTA guidelines reviewed by this thesis, only the HTA guideline of HIQA described the comparator choice for screening programmes. The above observation shows that the issue of comparator omission and the constraints of such evidence have not been well recognised by the research community and policy makers.

This thesis provides an accessible framework which is an ideal tool for addressing and communicating the importance of adequate comparator inclusion. The framework is editable by users and can illustrate the complete efficient frontier. Although HIQA recognises the difficulty of simulating a large number of screening strategies, it did not provide detailed guidance on how to deal with this. This thesis has the microsimulation framework simplified which can be easily comprehended and allows the exploration of a large number of strategies. From the exploration, this thesis provides guidance on how

CEAs should simulate their screening comparators efficiently with restricted computing capacity.

Other considerations in the decision-making

In addition to cost-effectiveness, other factors also influence decision-making in screening policy. Important considerations encompass capacity constraints in healthcare resources, budget impact, risk stratification, equity, adherence rate, and other concerns of decision-makers.

Regarding the capacity constraints in healthcare resources, this thesis suggests enhancing the healthcare system's capacity to support the optimal screening strategy. This is because the healthcare system can truly achieve maximum health benefits only when the ICER precisely aligns with the decision threshold. Limitations can stem from healthcare professionals, equipment, or financial budgets. This thesis proposes that addressing these resource limitations should be a separate discussion following the identification of cost-effective screening interventions. For instance, the budget impact analysis is an independent evaluation specifically assessing how the introduction of a new intervention impacts a healthcare system's overall budget. While in the short term, the healthcare system might not be able to provide sufficient screening services to match the optimal intensity, this thesis suggests that governments can temporarily prioritise screening for higher-risk individuals, which can lead to greater benefits.

Different decisions regarding which screening to offer can be made for each subgroup of patients with distinct characteristics. Subgroup analysis in CEAs facilitates examining costs and effectiveness for interventions applied to each subgroup. Each subgroup has its own unique optimal strategies, and decision-makers can consider applying different optimal strategies to corresponding subgroups. This represents the most "cost-effective" approach to allocating healthcare resources. Alternatively, applying the same screening strategy to the entire population can enhance equity in healthcare, ensuring equal access to preventive services for everyone. The optimal screening strategy for all groups may offer the highest benefit to all, some, or none of the subgroups (Drummond et al., 2015, p. 104). Consequently, a universal strategy might not be truly optimal but rather a compromise to promote equity or other social values.

Another factor to take into account is the adherence rate in screening. If policymakers are interested in understanding the effect of adherence rates, analysts should incorporate adherence considerations into CEAs of screening. A recommended approach is to conduct a sensitivity analysis, given the notable variability in observed screening adherence (Lopez-Olivo et al., 2020).

CEA serves as a decision-making tool in shaping healthcare policies, making it essential to consider additional factors that address the concerns or specific interests of decision-makers. If decision-makers have a clear intention of not adopting a particular strategy, there is a sound rationale to exclude it from consideration, thereby refining the scope of the study. Nevertheless, given that decision-makers might not be fully aware of the range of screening strategy possibilities, it falls upon analysts to inform them of these potential options. Analysts have the responsibility of providing reliable CEA evidence that can really address the research question, which commonly involves the broad objective of identifying the optimal screening strategy.

6. CONCLUSION

There is general consensus in economic evaluation theory regarding the need for incremental analysis. The reliability of CEA research depends not only on the transparency of reporting of results but also on the fundamentals of what strategies were compared. This thesis finds scope for improvement regarding the issue of strategy choice in many CEAs of CRC screening. CEAs of CRC screening that assess an insufficient range of alternatives likely offer biased results and mislead decision-makers. Biased evidence may cost society important population health gains. There appears scope for further research into further optimisation of population CRC screening through greater variation of screening intensities and combinations of modalities. The proposed checklist of this thesis may be a good basis for the development of a new screening simulation specific CEA checklist and can be beneficial when planning such research.

This thesis presents a simple microsimulation model of the cost-effectiveness of screening. This model is the first open-source DES CEA model of screening coded in R. It is specifically intended to overcome the constraints of the models typically applied in cancer screening, which are both large and not openly shared. In the initial application this thesis simulated thousands of screening strategies as an example to illustrate how the efficient frontier moves when parameter changes through a series of comparative statics. This permits a demonstration of face validity and is intended to aid modellers' understanding of screening cost-effectiveness. This thesis hopes this model will serve as a useful basis for methods research and as a teaching tool.

This thesis suggests CEAs of screening adopt an iterative approach to include sufficient cost-effective strategies. By reviewing the results of a selected subset of strategies, this thesis is able to anticipate the characteristics of the optimal strategy. Adjustment in the screening frequencies makes ICERs closer to the range of the policy-relevant portion of the decision threshold. This iterative approach solves the dilemma in the screening CEAs that limited simulation capacity possibly leads to the omission of the optimal strategy.

REFERENCES

Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: a tool for sequential decision making under uncertainty. *Med Decis Making*, *30*(4), 474-483. https://doi.org/10.1177/0272989x09353194

Alarid-Escudero, F., Krijkamp, E., Enns, E. A., Yang, A., Hunink, M. G. M., Pechlivanoglou, P., & Jalal, H. (2022). A Tutorial on Time-Dependent Cohort State-Transition Models in R Using a Cost-Effectiveness Analysis Example. *Med Decis Making*, 272989x221121747. https://doi.org/10.1177/0272989x221121747

Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*, *86*(4), 317-319. https://doi.org/10.2471/blt.07.050112

Arbyn, M., Raifu, A. O., Weiderpass, E., Bray, F., & Anttila, A. (2009). Trends of cervical cancer mortality in the member states of the European Union. *Eur J Cancer*, *45*(15), 2640-2648. https://doi.org/10.1016/j.ejca.2009.07.018

Areia, M., Fuccio, L., Hassan, C., Dekker, E., Dias-Pereira, A., & Dinis-Ribeiro, M. (2019). Cost-utility analysis of colonoscopy or faecal immunochemical test for population-based organised colorectal cancer screening. *United European Gastroenterol J*, *7*(1), 105-113. https://doi.org/10.1177/2050640618803196

Aronsson, M., Carlsson, P., Levin, L. Å., Hager, J., & Hultcrantz, R. (2017). Cost-effectiveness of high-sensitivity faecal immunochemical test and colonoscopy screening for colorectal cancer [Article]. *British Journal of Surgery*, *104*(8), 1078-1086. https://doi.org/10.1002/bjs.10536

Australian Government Department of Health. (2016). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 5.0)*. https://pbac.pbs.gov.au/

Azad, N. S., Leeds, I. L., Wanjau, W., Shin, E. J., & Padula, W. V. (2020). Cost-utility of colorectal cancer screening at 40 years old for average-risk patients. *Prev Med*, *133*, 106003. https://doi.org/10.1016/j.ypmed.2020.106003

Barouni, M., Larizadeh, M. H., Sabermahani, A., & Ghaderi, H. (2012). Markov's modeling for screening strategies for colorectal cancer. *Asian Pac J Cancer Prev*, *13*(10), 5125-5129. https://doi.org/10.7314/apjcp.2012.13.10.5125

Barzi, A., Lenz, H. J., Quinn, D. I., & Sadeghi, S. (2017). Comparative effectiveness of screening strategies for colorectal cancer. *Cancer*, *123*(9), 1516-1527. https://doi.org/10.1002/cncr.30518

Basu, P., Ponti, A., Anttila, A., Ronco, G., Senore, C., Vale, D. B., Segnan, N., Tomatis, M., Soerjomataram, I., Primic Žakelj, M., Dillner, J., Elfström, K. M., Lönnberg, S., & Sankaranarayanan, R. (2018). Status of implementation and organization of cancer screening in The European Union Member States-Summary results from the second European screening report. *Int J Cancer*, *142*(1), 44-56. https://doi.org/10.1002/ijc.31043

Berchi, C., Bouvier, V., Réaud, J. M., & Launoy, G. (2004). Cost-effectiveness analysis of two strategies for mass screening for colorectal cancer in France. *Health Econ*, *13*(3), 227-238. https://doi.org/10.1002/hec.819

Black, W. C. (1990). The CE plane: a graphic representation of cost-effectiveness. *Med Decis Making*, *10*(3), 212-214. https://doi.org/10.1177/0272989x9001000308

Blomström, P., Ekman, M., Lundqvist, C. B., Calvert, M. J., Freemantle, N., Lönnerholm, S., Wikström, G., & Jönsson, B. (2008). Cost effectiveness of

cardiac resynchronization therapy in the Nordic region: an analysis based on the CARE-HF trial. *Eur J Heart Fail*, *10*(9), 869-877. https://doi.org/10.1016/j.ejheart.2008.06.018

Boardman, A., Greenberg, D., Vining, A., & Weimer, D. (2014). *Cost-benefit analysis: Concepts and practice* (5th ed.). Harlow: Pearson Education Limited.

Bojke, L., Claxton, K., Sculpher, M., & Palmer, S. (2009). Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health*, 12(5), 739–749. https://doi.org/10.1111/j.1524-4733.2008.00502.x

Briggs, A., & Sculpher, M. (1998). An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics*, *13*(4), 397-409. https://doi.org/10.2165/00019053-199813040-00003

Briggs, A., Sculpher, M., & Claxton, K. (2006). *Decision modelling for health economic evaluation*. Oup Oxford.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J., & Paltiel, A. D. (2012). Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6. *Value Health*, *15*(6), 835-842. https://doi.org/10.1016/j.jval.2012.04.014

CADTH. (2017). *Guidelines for the economic evaluation of health technologies: Canada*. Ottawa: Canadian Agency for Drugs and Technologies in Health.

Campos, N. G., Castle, P. E., Wright, T. C., Jr., & Kim, J. J. (2015). Cervical cancer screening in low-resource settings: A cost-effectiveness framework for valuing tradeoffs between test performance and program coverage. *Int J Cancer*, *137*(9), 2208-2219. https://doi.org/10.1002/ijc.29594

Campos, N. G., Kim, J. J., Castle, P. E., Ortendahl, J. D., O'Shea, M., Diaz, M., & Goldie, S. J. (2012). Health and economic impact of HPV 16/18 vaccination and cervical cancer screening in Eastern Africa. *Int J Cancer*, *130*(11), 2672-2684. https://doi.org/10.1002/ijc.26269

Caro, J. J., Briggs, A. H., Siebert, U., & Kuntz, K. M. (2012). Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Med Decis Making*, *32*(5), 667-677. https://doi.org/10.1177/0272989x12454577

Chiou, C. F., Hay, J. W., Wallace, J. F., Bloom, B. S., Neumann, P. J., Sullivan, S. D., Yu, H. T., Keeler, E. B., Henning, J. M., & Ofman, J. J. (2003). Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care*, *41*(1), 32-44. https://doi.org/10.1097/00005650-200301000-00007

Cohen, J. T., Neumann, P. J., & Wong, J. B. (2017). A Call for Open-Source Cost-Effectiveness Analysis. *Ann Intern Med*, *167*(6), 432-433. https://doi.org/10.7326/m17-1153

Cohen, D. J., & Reynolds, M. R. (2008). Interpreting the Results of Cost-Effectiveness Studies. *Journal of the American College of Cardiology*, *52*(25), 2119-2126. https://doi.org/10.1016/j.jacc.2008.09.018

Coldman, A., Flanagan, W., Nadeau, C., Wolfson, M., Fitzgerald, N., Memon, S., Gauvreau, C., Miller, A., & Earle, C. (2017). Projected effect of fecal immunochemical test threshold for colorectal cancer screening on outcomes and costs for Canada using the OncoSim microsimulation model. *Journal of Cancer Policy*, *13*, 38-46. https://doi.org/https://doi.org/10.1016/j.jcpo.2017.07.004

Coretti, S., Ruggeri, M., Dibidino, R., Gitto, L., Marcellusi, A., Mennini, F. S., & Cicchetti, A. (2020). Economic evaluation of colorectal cancer screening

programs: Affordability for the health service. *J Med Screen*, *27*(4), 186-193. https://doi.org/10.1177/0969141319898732

Dan, Y. Y., Chuah, B. Y., Koh, D. C., & Yeoh, K. G. (2012). Screening based on risk for colorectal cancer is the most cost-effective approach. *Clin Gastroenterol Hepatol*, *10*(3), 266-271.e261-266. https://doi.org/10.1016/j.cgh.2011.11.011

Davis, S., Stevenson, M., Tappenden, P., & Wailoo, A. (2014). NICE Decision Support Unit Technical Support Documents. In *NICE DSU Technical Support Document 15: Cost-Effectiveness Modelling Using Patient-Level Simulation*. National Institute for Health and Care Excellence (NICE).

De Gelder, R., Heijnsdijk, E. A. M., Van Ravesteyn, N. T., Fracheboud, J., Draisma, G., & De Koning, H. J. (2011). Interpreting Overdiagnosis Estimates in Population-based Mammography Screening. *Epidemiologic Reviews*, *33*(1), 111-121. https://doi.org/10.1093/epirev/mxr009

de Kok, I. M., van Rosmalen, J., Dillner, J., Arbyn, M., Sasieni, P., Iftner, T., & van Ballegooijen, M. (2012). Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. *Bmj*, *344*, e670. https://doi.org/10.1136/bmj.e670

de Koning, H. J., Meza, R., Plevritis, S. K., ten Haaf, K., Munshi, V. N., Jeon, J., Erdogan, S. A., Kong, C. Y., Han, S. S., van Rosmalen, J., Choi, S. E., Pinsky, P. F., Berrington de Gonzalez, A., Berg, C. D., Black, W. C., Tammemägi, M. C., Hazelton, W. D., Feuer, E. J., & McMahon, P. M. (2014). Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*, *160*(5), 311-320. https://doi.org/10.7326/m13-2316

Dickinson, L. (1972). Evaluation of the effectiveness of cytologic screening for cervical cancer. 3. Cost-benefit analysis. *Mayo Clin Proc*, *47*(8), 550-555.

Dinh, T., Ladabaum, U., Alperin, P., Caldwell, C., Smith, R., & Levin, T. R. (2013). Health benefits and cost-effectiveness of a hybrid screening strategy for colorectal cancer. *Clin Gastroenterol Hepatol*, *11*(9), 1158-1166. https://doi.org/10.1016/j.cgh.2013.03.013

Drummond, M., & Sculpher, M. (2005). Common methodological flaws in economic evaluations. *Medical Care*, II5-II14.

Drummond, M. F., & Jefferson, T. O. (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *Bmj*, *313*(7052), 275-283. https://doi.org/10.1136/bmj.313.7052.275

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Oxford university press.

Dunlop, W. C. N., Mason, N., Kenworthy, J., & Akehurst, R. L. (2017). Benefits, Challenges and Potential Strategies of Open Source Health Economic Models. *Pharmacoeconomics*, *35*(1), 125-128. https://doi.org/10.1007/s40273-016-0479-8

Eddy, D. M. (1980). *Screening for cancer: Theory, analysis, and design*. Prentice Hall.

Eddy, D. M. (1981). The economics of cancer prevention and detection: getting more for less. Cancer, 47 (5 Suppl), 1200–1209. https://doi.org/10.1002/1097-0142(19810301)47:5+<1200::aid-cncr2820471325>3.0.co;2-6

Eddy, D. M. (1986). Secondary prevention of cancer: an overview. *Bulletin of the World Health Organization*, *64*(3), 421-429.

Eddy, D. M., Hollingworth, W., Caro, J. J., Tsevat, J., McDonald, K. M., & Wong, J. B. (2012). Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making*, *32*(5), 733-743. https://doi.org/10.1177/0272989x12454579

Enns, E. A., Cipriano, L. E., Simons, C. T., & Kong, C. Y. (2015). Identifying best-fitting inputs in health-economic model calibration: a Pareto frontier approach. *Med Decis Making*, *35*(2), 170-182. https://doi.org/10.1177/0272989x14528382

Etzioni, R., Gulati, R., Falcon, S., & Penson, D. F. (2008). Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: a surveillance modeling approach. *Med Decis Making*, *28*(3), 323-331. https://doi.org/10.1177/0272989x07312719

Evers, S., Goossens, M., de Vet, H., van Tulder, M., & Ament, A. (2005). Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care*, *21*(2), 240-245.

Fabbro, M., Hahn, K., Novaes, O., Ó'Grálaigh, M., & O'Mahony, J. F. (2022). Cost-Effectiveness Analyses of Lung Cancer Screening Using Low-Dose Computed Tomography: A Systematic Review Assessing Strategy Comparison and Risk Stratification. *Pharmacoecon Open*. https://doi.org/10.1007/s41669-022-00346-2

Fitch, K., Pyenson, B., Blumen, H., Weisman, T., & Small, A. (2015). The value of colonoscopic colorectal cancer screening of adults aged 50 to 64. *Am J Manag Care*, *21*(7), e430-438.

Flanagan, W. M., Le Petit, C., Berthelot, J. M., White, K. J., Coombs, B. A., & Jones-McLean, E. (2003). Potential impact of population-based colorectal cancer screening in Canada. *Chronic Dis Can*, *24*(4), 81-88.

Frazier, A. L., Colditz, G. A., Fuchs, C. S., & Kuntz, K. M. (2000). Cost-effectiveness of screening for colorectal cancer in the general population. *Jama*, *284*(15), 1954-1961. https://doi.org/10.1001/jama.284.15.1954

Frederix, G. W., van Hasselt, J. G., Severens, J. L., Hövels, A. M., Huitema, A. D., Raaijmakers, J. A., & Schellens, J. H. (2013). Development of a framework for cohort simulation in cost-effectiveness analyses using a multistep ordinary differential equation solver algorithm in R. *Med Decis Making*, *33*(6), 780-792. https://doi.org/10.1177/0272989x13476763

Frederix, G. W., Severens, J. L., & Hövels, A. M. (2015). Use of quality checklists and need for disease-specific guidance in economic evaluations: a meta-review. *Expert Rev Pharmacoecon Outcomes Res*, *15*(4), 675-685. https://doi.org/10.1586/14737167.2015.1069185

Gheysariyeha, F., Rahimi, F., Tabesh, E., Hemami, M. R., Adibi, P., & Rezayatmand, R. (2022). Cost-effectiveness of colorectal cancer screening strategies: A systematic review. *Eur J Cancer Care (Engl)*, e13673. https://doi.org/10.1111/ecc.13673

Goede, S. L., Rabeneck, L., van Ballegooijen, M., Zauber, A. G., Paszat, L. F., Hoch, J. S., Yong, J. H., Kroep, S., Tinmouth, J., & Lansdorp-Vogelaar, I. (2017). Harms, benefits and costs of fecal immunochemical testing versus guaiac fecal occult blood testing for colorectal cancer screening. *PLoS One*, *12*(3), e0172864. https://doi.org/10.1371/journal.pone.0172864

Goede, S. L., van Roon, A. H., Reijerink, J. C., van Vuuren, A. J., Lansdorp-Vogelaar, I., Habbema, J. D., Kuipers, E. J., van Leerdam, M. E., & van Ballegooijen, M. (2013). Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening. *Gut*, *62*(5), 727-734. https://doi.org/10.1136/gutjnl-2011-301917

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. Oxford university press.

Gray, A. M., Clarke, P. M., Wolstenholme, J. L., & Wordsworth, S. (2010). *Applied methods of cost-effectiveness analysis in healthcare* (Vol. 3). OUP Oxford.

Green, C., Handels, R., Gustavsson, A., Wimo, A., Winblad, B., Sköldunger, A., & Jönsson, L. (2019). Assessing cost-effectiveness of early intervention in Alzheimer's disease: An open-source modeling framework. *Alzheimers Dement*, *15*(10), 1309-1321. https://doi.org/10.1016/j.jalz.2019.05.004

Green, N., Lamrock, F., Naylor, N., Williams, J., & Briggs, A. (2022). Health Economic Evaluation Using Markov Models in R for Microsoft Excel Users: A Tutorial. *Pharmacoeconomics*. https://doi.org/10.1007/s40273-022-01199-7

Greuter, M. J., Berkhof, J., Fijneman, R. J., Demirel, E., Lew, J. B., Meijer, G. A., Stoker, J., & Coupé, V. M. (2016). The potential of imaging techniques as a screening tool for colorectal cancer: a cost-effectiveness analysis. *Br J Radiol*, *89*(1063), 20150910. https://doi.org/10.1259/bjr.20150910

Greuter, M. J., Carvalho, B., Wit, M., Dekker, E., Spaander, M. C., Meijer, G. A., Engeland, M. V., & Coupé, V. M. (2020). Can a biomarker triage test reduce colonoscopy burden in fecal immunochemical test screening? *J Comp Eff Res*, *9*(8), 563-571. https://doi.org/10.2217/cer-2019-0130

Greuter, M. J. E., de Klerk, C. M., Meijer, G. A., Dekker, E., & Coupé, V. M. H. (2017). Screening for Colorectal Cancer With Fecal Immunochemical Testing With and Without Postpolypectomy Surveillance Colonoscopy: A Cost-Effectiveness Analysis. *Ann Intern Med*, *167*(8), 544-554. https://doi.org/10.7326/m16-2891

Groot Koerkamp, B., Weinstein, M. C., Stijnen, T., Heijenbrok-Kal, M. H., & Hunink, M. G. (2010). Uncertainty and patient heterogeneity in medical decision models. *Med Decis Making*, *30*(2), 194-205. https://doi.org/10.1177/0272989x09342277

Grossman, D. C., Curry, S. J., Owens, D. K., Bibbins-Domingo, K., Caughey, A. B., Davidson, K. W., Doubeni, C. A., Ebell, M., Epling, J. W., Jr., Kemper, A. R., Krist, A. H., Kubik, M., Landefeld, C. S., Mangione, C. M., Silverstein, M., Simon, M. A., Siu, A. L., & Tseng, C. W. (2018). Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *Jama*, *319*(18), 1901-1913. https://doi.org/10.1001/jama.2018.3710

Gyrd-Hansen, D. (1998). Fecal occult blood tests. A cost-effectiveness analysis. *Int J Technol Assess Health Care*, *14*(2), 290-301. https://doi.org/10.1017/s0266462300012265

Gyrd-Hansen, D., Søgaard, J., & Kronborg, O. (1998). Colorectal cancer screening: efficiency and effectiveness. *Health Econ*, *7*(1), 9-20. https://doi.org/10.1002/(sici)1099-1050(199802)7:1<9::aid-hec304>3.0.co;2-h

Gulati, R., Gore, J. L., & Etzioni, R. (2013). Comparative effectiveness of alternative prostate-specific antigen--based prostate cancer screening strategies: model estimates of potential benefits and harms. *Ann Intern Med*, *158*(3), 145-153. https://doi.org/10.7326/0003-4819-158-3-201302050-00003

Gulati, R., Inoue, L., Katcher, J., Hazelton, W., & Etzioni, R. (2010). Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. *Biostatistics*, *11*(4), 707-719. https://doi.org/10.1093/biostatistics/kxq036

Hanly, P., Skally, M., Fenlon, H., & Sharp, L. (2012). Cost-effectiveness of computed tomography colonography in colorectal cancer screening: a systematic review. *Int J Technol Assess Health Care*, *28*(4), 415-423. https://doi.org/10.1017/s0266462312000542

Hassan, C., Benamouzig, R., Spada, C., Ponchon, T., Zullo, A., Saurin, J. C., & Costamagna, G. (2011). Cost effectiveness and projected national impact of colorectal cancer screening in France. *Endoscopy*, *43*(9), 780-793. https://doi.org/10.1055/s-0030-1256409

Hassan, C., Zullo, A., Winn, S., & Morini, S. (2008). Cost-effectiveness of capsule endoscopy in screening for colorectal cancer. *Endoscopy*, *40*(5), 414-421. https://doi.org/10.1055/s-2007-995565

Heitman, S. J., Hilsden, R. J., Au, F., Dowden, S., & Manns, B. J. (2010). Colorectal cancer screening for average-risk North Americans: an economic evaluation. *PLoS Med*, *7*(11), e1000370. https://doi.org/10.1371/journal.pmed.1000370

Heresbach, D., Chauvin, P., Grolier, J., & Josselin, J. M. (2010). Cost-effectiveness of colorectal cancer screening with computed tomography colonography or fecal blood tests. *Eur J Gastroenterol Hepatol*, *22*(11), 1372-1379. https://doi.org/10.1097/MEG.0b013e32833eaa71

HIQA. (2020). *Guidelines for the Economic Evaluation of Health Technologies in Ireland*. Health Information and Quality Authority.

Hoch, J. S., & Dewa, C. S. (2008). A Clinician's Guide to Correct Cost-Effectiveness Analysis: Think Incremental Not Average. *The Canadian Journal of Psychiatry*, *53*(4), 267-274. https://doi.org/10.1177/070674370805300408

Hollman, C., Paulden, M., Pechlivanoglou, P., & McCabe, C. (2017). A Comparison of Four Software Programs for Implementing Decision Analytic Cost-Effectiveness Models. *Pharmacoeconomics*, *35*(8), 817-830. https://doi.org/10.1007/s40273-017-0510-8

Hornik, K., & R Core Team, t. (2022). *The R FAQ*. https://CRAN.R-project.org/doc/FAQ/R-FAQ.html

Huang, W., Liu, G., Zhang, X., Fu, W., Zheng, S., Wu, Q., Liu, C., Liu, Y., Cai, S., & Huang, Y. (2014). Cost-effectiveness of colorectal cancer screening protocols in urban Chinese populations. *PLoS One*, *9*(10), e109150. https://doi.org/10.1371/journal.pone.0109150

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D., Augustovski, F., Briggs, A. H., Mauskopf, J., & Loder, E. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health*, *16*(2), 231-250. https://doi.org/10.1016/j.jval.2013.02.002

Imperiale, T. F., Kahi, C. J., & Rex, D. K. (2018). Lowering the Starting Age for Colorectal Cancer Screening to 45 Years: Who Will Come…and Should They? *Clinical Gastroenterology and Hepatology*, *16*(10), 1541-1544. https://doi.org/10.1016/j.cgh.2018.08.023

Incerti, D., Curtis, J. R., Shafrin, J., Lakdawalla, D. N., & Jansen, J. P. (2019). A Flexible Open-Source Decision Model for Value Assessment of Biologic Treatment for Rheumatoid Arthritis. *Pharmacoeconomics*, *37*(6), 829-843. https://doi.org/10.1007/s40273-018-00765-2

Incerti, D., & Jansen, J. P. (2021). Hesim: health economic simulation modeling and decision analysis. *arXiv preprint arXiv:2102.09437*.

Incerti, D., Thom, H., Baio, G., & Jansen, J. P. (2019). R You Still Using Excel? The Advantages of Modern Software Tools for Health Technology Assessment. *Value in Health*, *22*(5), 575-579. https://doi.org/10.1016/j.jval.2019.01.003

ISPOR. (2022). *Pharmacoeconomic Guidelines Around the World*. The Professional Society for Health Economics and Outcomes Research.

https://www.ispor.org/heor-resources/more-heor-resources/pharmacoeconomic-guidelines

Jahn, B., Sroczynski, G., Bundo, M., Mühlberger, N., Puntscher, S., Todorovic, J., Rochau, U., Oberaigner, W., Koffijberg, H., Fischer, T., Schiller-Fruehwirth, I., Öfner, D., Renner, F., Jonas, M., Hackl, M., Ferlitsch, M., & Siebert, U. (2019). Effectiveness, benefit harm and cost effectiveness of colorectal cancer screening in Austria. *BMC Gastroenterol*, *19*(1), 209. https://doi.org/10.1186/s12876-019-1121-y

Jalal, H., Pechlivanoglou, P., Krijkamp, E., Alarid-Escudero, F., Enns, E., & Hunink, M. G. M. (2017). An Overview of R in Health Decision Sciences. *Med Decis Making*, *37*(7), 735-746. https://doi.org/10.1177/0272989x16686559

Jansen, J. P., Incerti, D., & Linthicum, M. T. (2019). Developing Open-Source Models for the US Health System: Practical Experiences and Challenges to Date with the Open-Source Value Project. *Pharmacoeconomics*, *37*(11), 1313-1320. https://doi.org/10.1007/s40273-019-00827-z

Jeong, K. E., & Cairns, J. A. (2013). Review of economic evidence in the prevention and early detection of colorectal cancer. *Health Econ Rev*, *3*(1), 20. https://doi.org/10.1186/2191-1991-3-20

Jordan, M., Ghahramani, Z., & Saul, L. (1996). Hidden Markov decision trees. *Advances in neural information processing systems*, *9*.

Karlsson, A., Jauhiainen, A., Gulati, R., Eklund, M., Grönberg, H., Etzioni, R., & Clements, M. (2019). A natural history model for planning prostate cancer testing: Calibration and validation using Swedish registry data. *PLoS One*, *14*(2), e0211918. https://doi.org/10.1371/journal.pone.0211918

Karnon, J. (2003). Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics*, *12*(10), 837-848. https://doi.org/https://doi.org/10.1002/hec.770

Karnon, J., & Brown, J. (1998). Selecting a decision model for economic evaluation: a case study and review. *Health Care Management Science*, *1*(2), 133-140. https://doi.org/10.1023/A:1019090401655

Karnon, J., & Haji Ali Afzali, H. (2014). 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*, *32*(6), 547-558. https://doi.org/10.1007/s40273-014-0147-9

Karnon, J., Stahl, J., Brennan, A., Caro, J. J., Mar, J., & Möller, J. (2012). Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value Health*, *15*(6), 821-827. https://doi.org/10.1016/j.jval.2012.04.013

Keeney, E., Sanghera, S., Martin, R. M., Gulati, R., Wiklund, F., Walsh, E. I., Donovan, J. L., Hamdy, F., Neal, D. E., Lane, J. A., Turner, E. L., Thom, H., & Clements, M. S. (2022). Cost-Effectiveness Analysis of Prostate Cancer Screening in the UK: A Decision Model Analysis Based on the CAP Trial. *Pharmacoeconomics*. https://doi.org/10.1007/s40273-022-01191-1

Khalili, F., Najafi, B., Mansour-Ghanaei, F., Yousefi, M., Abdollahzad, H., & Motlagh, A. (2020). Cost-Effectiveness Analysis of Colorectal Cancer Screening: A Systematic Review. *Risk Manag Healthc Policy*, *13*, 1499-1512. https://doi.org/10.2147/rmhp.S262171

Khandker, R. K., Dulski, J. D., Kilpatrick, J. B., Ellis, R. P., Mitchell, J. B., & Baine, W. B. (2000). A decision model and cost-effectiveness analysis of colorectal cancer

screening and surveillance guidelines for average-risk adults. *Int J Technol Assess Health Care*, *16*(3), 799-810. https://doi.org/10.1017/s0266462300102077

Knudsen, A. B., Lansdorp-Vogelaar, I., Rutter, C. M., Savarino, J. E., van Ballegooijen, M., Kuntz, K. M., & Zauber, A. G. (2010). Cost-effectiveness of computed tomographic colonography screening for colorectal cancer in the medicare population. *J Natl Cancer Inst*, *102*(16), 1238-1252. https://doi.org/10.1093/jnci/djq242

Knudsen, A. B., McMahon, P. M., & Gazelle, G. S. (2007). Use of modeling to evaluate the cost-effectiveness of cancer screening programs. *J Clin Oncol*, *25*(2), 203-208. https://doi.org/10.1200/jco.2006.07.9202

Koffijberg, H., Degeling, K., Ijzerman, M.J., Coupé, V.M.H., Greuter, M.J.E. (2021). Using metamodeling to identify the optimal strategy for colorectal cancer screening. Value in Health 24, 206–215.. https://doi.org/10.1016/j.jval.2020.08.2099

Krijkamp, E. M., Alarid-Escudero, F., Enns, E. A., Jalal, H. J., Hunink, M. G. M., & Pechlivanoglou, P. (2018). Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making*, *38*(3), 400-422. https://doi.org/10.1177/0272989x18754513

Krijkamp, E. M., Alarid-Escudero, F., Enns, E. A., Pechlivanoglou, P., Hunink, M. G. M., Yang, A., & Jalal, H. J. (2020). A Multidimensional Array Representation of State-Transition Model Dynamics. *Med Decis Making*, *40*(2), 242-248. https://doi.org/10.1177/0272989x19893973

Kriza, C., Emmert, M., Wahlster, P., Niederlander, C., & Kolominsky-Rabas, P. (2013). An international review of the main cost-effectiveness drivers of virtual colonography versus conventional colonoscopy for colorectal cancer screening: is the tide changing due to adherence? *Eur J Radiol*, *82*(11), e629-636. https://doi.org/10.1016/j.ejrad.2013.07.019

Kuntz, K. M., Russell, L. B., Owens, D. K., Sanders, G. D., Trikalinos, T. A., & Salomon, J. A. (2016). Decision Models in Cost-Effectiveness Analysis. In P. J. Neumann, T. G. Ganiats, L. B. Russell, G. D. Sanders, & J. E. Siegel (Eds.), *Cost-effectiveness in health and medicine* (pp. 105-136). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190492939.003.0005

Lăără, E., Day, N. E., & Hakama, M. (1987). Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet*, *1*(8544), 1247-1249. https://doi.org/10.1016/s0140-6736(87)92695-x

Ladabaum, U., Allen, J., Wandell, M., & Ramsey, S. (2013). Colorectal cancer screening with blood-based biomarkers: cost-effectiveness of methylated septin 9 DNA versus current strategies. *Cancer Epidemiol Biomarkers Prev*, *22*(9), 1567-1576. https://doi.org/10.1158/1055-9965.Epi-13-0204

Ladabaum, U., Alvarez-Osorio, L., Rösch, T., & Brueggenjuergen, B. (2014). Cost-effectiveness of colorectal cancer screening in Germany: current endoscopic and fecal testing strategies versus plasma methylated Septin 9 DNA. *Endosc Int Open*, *2*(2), E96-e104. https://doi.org/10.1055/s-0034-1377182

Ladabaum, U., & Mannalithara, A. (2016). Comparative Effectiveness and Cost Effectiveness of a Multitarget Stool DNA Test to Screen for Colorectal Neoplasia. *Gastroenterology*, *151*(3), 427-439.e426. https://doi.org/10.1053/j.gastro.2016.06.003

Ladabaum, U., Mannalithara, A., Meester, R. G. S., Gupta, S., & Schoen, R. E. (2019). Cost-Effectiveness and National Effects of Initiating Colorectal Cancer

Screening for Average-Risk Persons at Age 45 Years Instead of 50 Years. *Gastroenterology*, *157*(1), 137-148. https://doi.org/10.1053/j.gastro.2019.03.023

Ladabaum, U., Song, K., & Fendrick, A. M. (2004). Colorectal neoplasia screening with virtual colonoscopy: when, at what cost, and with what national impact? *Clin Gastroenterol Hepatol*, *2*(7), 554-563. https://doi.org/10.1016/s1542-3565(04)00247-2

Lansdorp-Vogelaar, I., Knudsen, A. B., & Brenner, H. (2011). Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev*, *33*(1), 88-100. https://doi.org/10.1093/epirev/mxr004

Lansdorp-Vogelaar, I., Kuntz, K. M., Knudsen, A. B., Wilschut, J. A., Zauber, A. G., & van Ballegooijen, M. (2010). Stool DNA testing to screen for colorectal cancer in the Medicare population: a cost-effectiveness analysis. *Ann Intern Med*, *153*(6), 368-377. https://doi.org/10.7326/0003-4819-153-6-201009210-00004

Lee, D., Muston, D., Sweet, A., Cunningham, C., Slater, A., & Lock, K. (2010). Cost effectiveness of CT colonography for UK NHS colorectal cancer screening of asymptomatic adults aged 60-69 years. *Appl Health Econ Health Policy*, *8*(3), 141-154. https://doi.org/10.2165/11535650-000000000-00000

Lee, S. J., & Zelen, M. (2008). Mortality modeling of early detection programs. *Biometrics*, *64*(2), 386-395. https://doi.org/10.1111/j.1541-0420.2007.00893.x

Lejeune, C., Arveux, P., Dancourt, V., Béjean, S., Bonithon-Kopp, C., & Faivre, J. (2004). Cost-effectiveness analysis of fecal occult blood screening for colorectal cancer. *Int J Technol Assess Health Care*, *20*(4), 434-439. https://doi.org/10.1017/s0266462304001321

Lejeune, C., Dancourt, V., Arveux, P., Bonithon-Kopp, C., & Faivre, J. (2010). Cost-effectiveness of screening for colorectal cancer in France using a guaiac test versus an immunochemical test. *Int J Technol Assess Health Care*, *26*(1), 40-47. https://doi.org/10.1017/s026646230999078x

Lejeune, C., Le Gleut, K., Cottet, V., Galimard, C., Durand, G., Dancourt, V., & Faivre, J. (2014). The cost-effectiveness of immunochemical tests for colorectal cancer screening. *Dig Liver Dis*, *46*(1), 76-81. https://doi.org/10.1016/j.dld.2013.07.018

Lew, J. B., St John, D. J. B., Macrae, F. A., Emery, J. D., Ee, H. C., Jenkins, M. A., He, E., Grogan, P., Caruana, M., Greuter, M. J. E., Coupé, V. M. H., & Canfell, K. (2018). Benefits, Harms, and Cost-Effectiveness of Potential Age Extensions to the National Bowel Cancer Screening Program in Australia. *Cancer Epidemiol Biomarkers Prev*, *27*(12), 1450-1461. https://doi.org/10.1158/1055-9965.Epi-18-0128

Lew, J. B., St John, D. J. B., Macrae, F. A., Emery, J. D., Ee, H. C., Jenkins, M. A., He, E., Grogan, P., Caruana, M., Sarfati, D., Greuter, M. J. E., Coupé, V. M. H., & Canfell, K. (2018). Evaluation of the benefits, harms and cost-effectiveness of potential alternatives to iFOBT testing for colorectal cancer screening in Australia. *Int J Cancer*, *143*(2), 269-282. https://doi.org/10.1002/ijc.31314

Lew, J. B., St John, D. J. B., Xu, X. M., Greuter, M. J. E., Caruana, M., Cenin, D. R., He, E., Saville, M., Grogan, P., Coupé, V. M. H., & Canfell, K. (2017). Long-term evaluation of benefits, harms, and cost-effectiveness of the National Bowel Cancer Screening Program in Australia: a modelling study. *Lancet Public Health*, *2*(7), e331-e340. https://doi.org/10.1016/s2468-2667(17)30105-6

Lopez-Olivo, M.A., Maki, K.G., Choi, N.J., Hoffman, R.M., Shih, Y.-C.T., Lowenstein, L.M., Hicklen, R.S., Volk, R.J. (2020). Patient adherence to screening for lung cancer in the US. JAMA Network Open 3, e2025102.. https://doi.org/10.1001/jamanetworkopen.2020.25102

Macafee, D. A., Waller, M., Whynes, D. K., Moss, S., & Scholefield, J. H. (2008). Population screening for colorectal cancer: the implications of an ageing population. *Br J Cancer*, *99*(12), 1991-2000. https://doi.org/10.1038/sj.bjc.6604788

Mandelblatt, J., Fraback, D., Weinstein, M., Russell, L., Gold, M., & Hadorn, D. (1996). Assessing the effectiveness of health interventions. In *Cost-effectiveness in health and medicine* (pp. 135-175).

Mannucci, A., Zuppardo, R. A., Rosati, R., Leo, M. D., Perea, J., & Cavestro, G. M. (2019). Colorectal cancer screening from 45 years of age: Thesis, antithesis and synthesis. *World Journal of Gastroenterology*, *25*(21), 2565-2580. https://doi.org/10.3748/wjg.v25.i21.2565

Marshall, D. A., Burgos-Liz, L., MJ, I. J., Osgood, N. D., Padula, W. V., Higashi, M. K., Wong, P. K., Pasupathy, K. S., & Crown, W. (2015). Applying dynamic simulation modeling methods in health care delivery research-the SIMULATE checklist: report of the ISPOR simulation modeling emerging good practices task force. *Value Health*, *18*(1), 5-16. https://doi.org/10.1016/j.jval.2014.12.001

McKenna, C., Hawkins, N., Claxton, K., McDaid, C., Suekarran, S., Light, K., Chester, M., Cleland, J. G., Woolacott, N., & Sculpher, M. (2010). Cost-effectiveness of enhanced external counterpulsation (EECP) for the treatment of stable angina in the United Kingdom. *Int J Technol Assess Health Care*, *26*(2), 175-182. https://doi.org/10.1017/s0266462310000073

McLeod, M., Kvizhinadze, G., Boyd, M., Barendregt, J., Sarfati, D., Wilson, N., & Blakely, T. (2017). Colorectal Cancer Screening: How Health Gains and Cost-Effectiveness Vary by Ethnic Group, the Impact on Health Inequalities, and the Optimal Age Range to Screen. *Cancer Epidemiol Biomarkers Prev*, *26*(9), 1391-1400. https://doi.org/10.1158/1055-9965.Epi-17-0150

Melnitchouk, N., Soeteman, D. I., Davids, J. S., Fields, A., Cohen, J., Noubary, F., Lukashenko, A., Kolesnik, O. O., & Freund, K. M. (2018). Cost-effectiveness of colorectal cancer screening in Ukraine. *Cost Eff Resour Alloc*, *16*, 20. https://doi.org/10.1186/s12962-018-0104-0

Mendivil, J., Appierto, M., Aceituno, S., Comas, M., & Rué, M. (2019). Economic evaluations of screening strategies for the early detection of colorectal cancer in the average-risk population: A systematic literature review. *PLoS One*, *14*(12), e0227251. https://doi.org/10.1371/journal.pone.0227251

Menn, P., & Holle, R. (2009). Comparing three software tools for implementing markov models for health economic evaluations. *Pharmacoeconomics*, *27*(9), 745-753. https://doi.org/10.2165/11313760-000000000-00000

Mezei, A. K., Armstrong, H. L., Pedersen, H. N., Campos, N. G., Mitchell, S. M., Sekikubo, M., Byamugisha, J. K., Kim, J. J., Bryan, S., & Ogilvie, G. S. (2017). Cost-effectiveness of cervical cancer screening methods in low- and middle-income countries: A systematic review. *Int J Cancer*, *141*(3), 437-446. https://doi.org/10.1002/ijc.30695

Min, C., Xue, M., Haotian, F., Jialian, L., & Lingli, Z. (2021). An overview of the characteristics and quality assessment criteria in systematic review of pharmacoeconomics. *PLoS One*, *16*(2), e0246080. https://doi.org/10.1371/journal.pone.0246080

Murphy, J., Halloran, S., & Gray, A. (2017). Cost-effectiveness of the faecal immunochemical test at a range of positivity thresholds compared with the guaiac faecal occult blood test in the NHS Bowel Cancer Screening Programme in

England. *BMJ Open*, *7*(10), e017186. https://doi.org/10.1136/bmjopen-2017-017186

Naber, S. K., Knudsen, A. B., Zauber, A. G., Rutter, C. M., Fischer, S. E., Pabiniak, C. J., Soto, B., Kuntz, K. M., & Lansdorp-Vogelaar, I. (2019). Cost-effectiveness of a multitarget stool DNA test for colorectal cancer screening of Medicare beneficiaries. *PLoS One*, *14*(9), e0220234. https://doi.org/10.1371/journal.pone.0220234

Naber, S. K., Matthijsse, S. M., Rozemeijer, K., Penning, C., De Kok, I. M. C. M., & Van Ballegooijen, M. (2016). Cervical Cancer Screening in Partly HPV Vaccinated Cohorts – A Cost-Effectiveness Analysis. *PLoS One*, *11*(1), e0145548. https://doi.org/10.1371/journal.pone.0145548

Neuhauser, D., & Lewicki, A. M. (1975). What do we gain from the sixth stool guaiac? *N Engl J Med*, *293*(5), 226-228. https://doi.org/10.1056/nejm197507312930504

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2016). *Cost-Effectiveness in Health and Medicine*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190492939.001.0001

Ngan, T. T., Browne, S., Goodwin, M., Van Minh, H., Donnelly, M., & O'Neill, C. (2022). Cost-effectiveness of clinical breast examination screening programme among HER2-positive breast cancer patients: a modelling study. *Breast Cancer*. https://doi.org/10.1007/s12282-022-01398-2

NICE. (2022). *Guide to the methods of technology appraisal*. National Institute for Health and Care Excellence. https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation

O'Leary, B. A., Olynyk, J. K., Neville, A. M., & Platell, C. F. (2004). Cost-effectiveness of colorectal cancer screening: comparison of community-based flexible sigmoidoscopy with fecal occult blood testing and colonoscopy. *J Gastroenterol Hepatol*, *19*(1), 38-47. https://doi.org/10.1111/j.1440-1746.2004.03177.x

O'Mahony, J. F. (2014). *Revisiting the simulation evidence for the cost-effectiveness of breast cancer screening of average-risk women* 15th Biennial European Meeting of the Society for Medical Decision Making, Antwerp. https://journals.sagepub.com/doi/abs/10.1177/0272989X14547195

O'Mahony, J. F. (2019). *How Complete is the Evidence that Biennial Mammography in Average Risk Women is Cost-Effective? Revisiting the Literature* International Cancer Screening Conference, Rotterdam.

O'Mahony, J. F., Naber, S. K., Normand, C., Sharp, L., O'Leary, J. J., & De Kok, I. M. C. M. (2015). Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. *Value in Health*, *18*(8), 1138-1151. https://doi.org/10.1016/j.jval.2015.09.2939

Owens, D. K., Siegel, J. E., Sculpher, M. J., & Salomon, J. A. (2016). Designing a Cost-Effectiveness Analysis. In P. J. Neumann, T. G. Ganiats, L. B. Russell, G. D. Sanders, & J. E. Siegel (Eds.), *Cost-effectiveness in health and medicine* (pp. 75-104). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190492939.003.0004

Parekh, M., Fendrick, A. M., & Ladabaum, U. (2008). As tests evolve and costs of cancer care rise: reappraising stool-based screening for colorectal neoplasia. *Aliment Pharmacol Ther*, *27*(8), 697-712. https://doi.org/10.1111/j.1365-2036.2008.03632.x

Patel, S. S., & Kilgore, M. L. (2015). Cost Effectiveness of Colorectal Cancer Screening Strategies. *Cancer Control*, *22*(2), 248-258. https://doi.org/10.1177/107327481502200219

Peterse, E. F. P., Meester, R. G. S., de Jonge, L., Omidvari, A. H., Alarid-Escudero, F., Knudsen, A. B., Zauber, A. G., & Lansdorp-Vogelaar, I. (2021). Comparing the Cost-Effectiveness of Innovative Colorectal Cancer Screening Tests. *J Natl Cancer Inst*, *113*(2), 154-161. https://doi.org/10.1093/jnci/djaa103

Philips, Z., Bojke, L., Sculpher, M., Claxton, K., & Golder, S. (2006). Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*, *24*(4), 355-371. https://doi.org/10.2165/00019053-200624040-00006

Phisalprapa, P., Supakankunti, S., & Chaiyakunapruk, N. (2019). Cost-effectiveness and budget impact analyses of colorectal cancer screenings in a low- and middle-income country: example from Thailand. *J Med Econ*, *22*(12), 1351-1361. https://doi.org/10.1080/13696998.2019.1674065

Pickhardt, P. J., Hassan, C., Laghi, A., Zullo, A., Kim, D. H., & Morini, S. (2007). Cost-effectiveness of colorectal cancer screening with computed tomography colonography. *Cancer*, *109*(11), 2213-2221. https://doi.org/https://doi.org/10.1002/cncr.22668

Pignone, M., Saha, S., Hoerger, T., & Mandelblatt, J. (2002). Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med*, *137*(2), 96-104. https://doi.org/10.7326/0003-4819-137-2-200207160-00007

Pignone, M. P., Flitcroft, K. L., Howard, K., Trevena, L. J., Salkeld, G. P., & St John, D. J. (2011). Costs and cost-effectiveness of full implementation of a biennial faecal occult blood test screening program for bowel cancer in Australia. *Med J Aust*, *194*(4), 180-185. https://doi.org/10.5694/j.1326-5377.2011.tb03766.x

Ponti, A., Anttila, A., Ronco, G., & Senore, C. (2017). Cancer screening in the European Union (2017). Report on the implementation of the council recommendation on cancer screening. *Cancer screening in the European Union (2017). Report on the implementation of the council recommendation on cancer screening.*

Poole, C., Agrawal, S., & Currie, C. J. (2007). Let cost effectiveness models be open to scrutiny. *Bmj*, *335*(7623), 735. https://doi.org/10.1136/bmj.39360.379664.BE

Prakash, M. K., Lang, B., Heinrich, H., Valli, P. V., Bauerfeind, P., Sonnenberg, A., Beerenwinkel, N., & Misselwitz, B. (2017). CMOST: an open-source framework for the microsimulation of colorectal cancer screening strategies. *BMC Med Inform Decis Mak*, *17*(1), 80. https://doi.org/10.1186/s12911-017-0458-9

Prosser, L. A., Neumann, P. J., Sanders, G. D., & Siegel, J. E. (2016). Reporting cost-effectiveness analyses. In *Cost-Effectiveness in Health and Medicine.* (pp. 343-368). Oxford University Press.

Ran, T., Cheng, C. Y., Misselwitz, B., Brenner, H., Ubels, J., & Schlander, M. (2019). Cost-Effectiveness of Colorectal Cancer Screening Strategies-A Systematic Review. *Clin Gastroenterol Hepatol*, *17*(10), 1969-1981.e1915. https://doi.org/10.1016/j.cgh.2019.01.014

Rashidian, A., Barfar, E., Hosseini, H., Nosratnejad, S., & Barooti, E. (2013). Cost effectiveness of breast cancer screening using mammography; a systematic review. *Iran J Public Health*, *42*(4), 347-357.

Regge, D., Hassan, C., Pickhardt, P. J., Laghi, A., Zullo, A., Kim, D. H., Iafrate, F., & Morini, S. (2009). Impact of computer-aided detection on the cost-effectiveness

of CT colonography. *Radiology*, *250*(2), 488-497. https://doi.org/10.1148/radiol.2502080685

Røe, O. D. (2020). Democratic and ethical problem of lung cancer screening: exclusion of true high-risk populations. Can it be fixed? Yes. *BMJ Open Respir Res*, *7*(1). https://doi.org/10.1136/bmjresp-2020-000811

Rutter, C. M., & Savarino, J. E. (2010). An evidence-based microsimulation model for colorectal cancer: validation and application. *Cancer Epidemiol Biomarkers Prev*, *19*(8), 1992-2002. https://doi.org/10.1158/1055-9965.Epi-09-0954

Salkeld, G., Young, G., Irwig, L., Haas, M., & Glasziou, P. (1996). Cost-effectiveness analysis of screening by faecal occult blood testing for colorectal cancer in Australia. *Aust N Z J Public Health*, *20*(2), 138-143. https://doi.org/10.1111/j.1753-6405.1996.tb01807.x

Sanghera, S., Coast, J., Martin, R. M., Donovan, J. L., & Mohiuddin, S. (2018). Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models. *BMC Cancer*, *18*(1), 84. https://doi.org/10.1186/s12885-017-3974-1

Saslow, D., Runowicz, C. D., Solomon, D., Moscicki, A. B., Smith, R. A., Eyre, H. J., & Cohen, C. (2002). American Cancer Society guideline for the early detection of cervical neoplasia and cancer. *CA Cancer J Clin*, *52*(6), 342-362. https://doi.org/10.3322/canjclin.52.6.342

Schiller-Frühwirth, I. C., Jahn, B., Arvandi, M., & Siebert, U. (2017). Cost-Effectiveness Models in Breast Cancer Screening in the General Population: A Systematic Review. *Appl Health Econ Health Policy*, *15*(3), 333-351. https://doi.org/10.1007/s40258-017-0312-3

Schramm, W., Sailer, F., Pobiruchin, M., & Weiss, C. (2016). PROSIT Open Source Disease Models for Diabetes Mellitus. *Stud Health Technol Inform*, *226*, 115-118.

Schreuders, E. H., Ruco, A., Rabeneck, L., Schoen, R. E., Sung, J. J., Young, G. P., & Kuipers, E. J. (2015). Colorectal cancer screening: a global overview of existing programmes. *Gut*, *64*(10), 1637-1649. https://doi.org/10.1136/gutjnl-2014-309086

Schweitzer, S. O. (1974). Cost effectiveness of early detection of disease. *Health Serv Res*, *9*(1), 22-32.

Sekiguchi, M., Igarashi, A., Matsuda, T., Matsumoto, M., Sakamoto, T., Nakajima, T., Kakugawa, Y., Yamamoto, S., Saito, H., & Saito, Y. (2016). Optimal use of colonoscopy and fecal immunochemical test for population-based colorectal cancer screening: a cost-effectiveness analysis using Japanese data. *Jpn J Clin Oncol*, *46*(2), 116-125. https://doi.org/10.1093/jjco/hyv186

Sekiguchi, M., Igarashi, A., Sakamoto, T., Saito, Y., Esaki, M., & Matsuda, T. (2020). Cost-effectiveness analysis of colorectal cancer screening using colonoscopy, fecal immunochemical test, and risk score. *J Gastroenterol Hepatol*, *35*(9), 1555-1561. https://doi.org/10.1111/jgh.15033

Senore, C., Hassan, C., Regge, D., Pagano, E., Iussich, G., Correale, L., & Segnan, N. (2019). Cost-effectiveness of colorectal cancer screening programmes using sigmoidoscopy and immunochemical faecal occult blood test. *J Med Screen*, *26*(2), 76-83. https://doi.org/10.1177/0969141318789710

Sharaf, R. N., & Ladabaum, U. (2013). Comparative effectiveness and cost-effectiveness of screening colonoscopy vs. sigmoidoscopy and alternative strategies. *Am J Gastroenterol*, *108*(1), 120-132. https://doi.org/10.1038/ajg.2012.380

Sharp, L., Tilson, L., Whyte, S., O'Ceilleachair, A., Walsh, C., Usher, C., Tappenden, P., Chilcott, J., Staines, A., Barry, M., & Comber, H. (2012). Cost-effectiveness of population-based screening for colorectal cancer: a comparison of guaiac-based faecal occult blood testing, faecal immunochemical testing and flexible sigmoidoscopy. *Br J Cancer*, *106*(5), 805-816. https://doi.org/10.1038/bjc.2011.580

Shi, J. F., Canfell, K., Lew, J. B., Zhao, F. H., Legood, R., Ning, Y., Simonella, L., Ma, L., Kang, Y. J., Zhang, Y. Z., Smith, M. A., Chen, J. F., Feng, X. X., & Qiao, Y. L. (2011). Evaluation of primary HPV-DNA testing in relation to visual inspection methods for cervical cancer screening in rural China: an epidemiologic and cost-effectiveness modelling study. *BMC Cancer*, *11*, 239. https://doi.org/10.1186/1471-2407-11-239

Shimbo, T., Glick, H. A., & Eisenberg, J. M. (1994). Cost-effectiveness analysis of strategies for colorectal cancer screening in Japan. *Int J Technol Assess Health Care*, *10*(3), 359-375. https://doi.org/10.1017/s0266462300006607 Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D. J., & Kuntz, K. M. (2012). State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Value Health*, *15*(6), 812-820. https://doi.org/10.1016/j.jval.2012.06.014

Siegel, J. E., Weinstein, M., & Torrance, G. (1996). Reporting cost-effectiveness studies and results. In *Cost-effectiveness in health and medicine* (pp. 276-303).

Skally, M., Hanly, P., & Sharp, L. (2013). Cost effectiveness of fecal DNA screening for colorectal cancer: a systematic review and quality appraisal of the literature. *Appl Health Econ Health Policy*, *11*(3), 181-192. https://doi.org/10.1007/s40258-013-0010-8

Simpson, K. N., Strassburger, A., Jones, W. J., Dietz, B., & Rajagopalan, R. (2009). Comparison of Markov model and discrete-event simulation techniques for HIV. *Pharmacoeconomics*, *27*(2), 159-165. https://doi.org/10.2165/00019053-200927020-00006

Srivastava, S., Koay, E. J., Borowsky, A. D., De Marzo, A. M., Ghosh, S., Wagner, P. D., & Kramer, B. S. (2019). Cancer overdiagnosis: a biological challenge and clinical dilemma. *Nat Rev Cancer*, *19*(6), 349-358. https://doi.org/10.1038/s41568-019-0142-8

Smith, R. A., Saslow, D., Sawyer, K. A., Burke, W., Costanza, M. E., Evans, W. P., 3rd, Foster, R. S., Jr., Hendrick, E., Eyre, H. J., & Sener, S. (2003). American Cancer Society guidelines for breast cancer screening: update 2003. *CA Cancer J Clin*, *53*(3), 141-169. https://doi.org/10.3322/canjclin.53.3.141

Song, K., Fendrick, A. M., & Ladabaum, U. (2004). Fecal DNA testing compared with conventional colorectal cancer screening methods: a decision analysis. *Gastroenterology*, *126*(5), 1270-1279. https://doi.org/10.1053/j.gastro.2004.02.016

Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making: a practical guide. *Med Decis Making*, *13*(4), 322-338. https://doi.org/10.1177/0272989x9301300409

Sonnenberg, A., & Delco, F. (2002). Cost-effectiveness of a single colonoscopy in screening for colorectal cancer. *Arch Intern Med*, *162*(2), 163-168.

Sonnenberg, A., Delcò, F., & Bauerfeind, P. (1999). Is virtual colonoscopy a cost-effective option to screen for colorectal cancer? *Am J Gastroenterol*, *94*(8), 2268-2274. https://doi.org/10.1111/j.1572-0241.1999.01304.x

Sonnenberg, A., Delco, F., & Inadomi, J. M. (2000). Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Ann Intern Med*, *133*(8), 573-584. https://doi.org/10.7326/0003-4819-133-8-200010170-00007

Standfield, L., Comans, T., & Scuffham, P. (2014). Markov modeling and discrete event simulation in health care: a systematic comparison. *Int J Technol Assess Health Care*, *30*(2), 165-172. https://doi.org/10.1017/s0266462314000117

Stinnett, A. A., Mittleman, M. A., Weinstein, M. C., Kuntz, K. M., Cohen, D. J., Williams, L. W., Goldman, P. A., Staiger, D. O., Hunink, M. G. M., Tsevat, J., Tosteson, A. N. A., & Goldman, L. (1996). The cost-effectiveness of dietary and pharmacologic therapy for cholesterol reduction in adults. In *Cost-effectiveness in health and medicine* (pp. 349-391).

Stoler, M. H., Baker, E., Boyle, S., Aslam, S., Ridder, R., Huh, W. K., & Wright, T. C. (2020). Approaches to triage optimization in HPV primary screening: Extended genotyping and p16/Ki-67 dual-stained cytology—Retrospective insights from ATHENA. *International Journal of Cancer*, *146*(9), 2599-2607. https://doi.org/10.1002/ijc.32669

Subramanian, S., Bobashev, G., & Morris, R. J. (2009). Modeling the cost-effectiveness of colorectal cancer screening: policy guidance based on patient preferences and compliance. *Cancer Epidemiol Biomarkers Prev*, *18*(7), 1971-1978. https://doi.org/10.1158/1055-9965.Epi-09-0083

Sullivan, W., Hirst, M., Beard, S., Gladwell, D., Fagnani, F., López Bastida, J., Phillips, C., & Dunlop, W. C. (2016). Economic evaluation in chronic pain: a systematic review and de novo flexible economic model. *Eur J Health Econ*, *17*(6), 755-770. https://doi.org/10.1007/s10198-015-0720-y

Tafazzoli, A., Roberts, S., Klein, R., Ness, R., & Dittus, R. (2009). Probabilistic cost-effectiveness comparison of screening strategies for colorectal cancer. *ACM Transactions on Modeling and Computer Simulation*, *19*(2), 1-29. https://doi.org/10.1145/1502787.1502789

Tappenden, P., Chilcott, J., Eggington, S., Patnick, J., Sakai, H., & Karnon, J. (2007). Option appraisal of population-based colorectal cancer screening programmes in England. *Gut*, *56*(5), 677-684. https://doi.org/10.1136/gut.2006.095109

Telford, J. J., Levy, A. R., Sambrook, J. C., Zou, D., & Enns, R. A. (2010). The cost-effectiveness of screening for colorectal cancer. *Cmaj*, *182*(12), 1307-1313. https://doi.org/10.1503/cmaj.090845

Ten Haaf, K., Tammemägi, M. C., Bondy, S. J., Van Der Aalst, C. M., Gu, S., McGregor, S. E., Nicholas, G., De Koning, H. J., & Paszat, L. F. (2017). Performance and Cost-Effectiveness of Computed Tomography Lung Cancer Screening Scenarios in a Population-Based Setting: A Microsimulation Modeling Analysis in Ontario, Canada. *PLOS Medicine*, *14*(2), e1002225. https://doi.org/10.1371/journal.pmed.1002225

Ten Haaf, K., van der Aalst, C. M., de Koning, H. J., Kaaks, R., & Tammemägi, M. C. (2021). Personalising lung cancer screening: An overview of risk-stratification opportunities and challenges. *International Journal of Cancer*, *149*(2), 250-263.

Thankappan, K., Subramanian, S., Balasubramanian, D., Kuriakose, M. A., Sankaranarayanan, R., & Iyer, S. (2021). Cost-effectiveness of oral cancer screening approaches by visual examination: Systematic review. *Head Neck*, *43*(11), 3646-3661. https://doi.org/10.1002/hed.26816

Tjalma, W. (2018). HPV negative cervical cancers and primary HPV screening. *Facts Views Vis Obgyn*, *10*(2), 107-113.

Torrance, G., Siegel, J. E., & Luce, B. (1996). Framing and designing the cost-effectiveness analysis. In *Cost-effectiveness in health and medicine* (pp. 54-81).

Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A utility maximization model for evaluation of health care programs. *Health services research*, *7*(2), 118.

Tosh, J., & Wailoo, A. (2008). NICE Decision Support Unit Methods Development. In *Review of Software for Decision Modelling*. National Institute for Health and Care Excellence (NICE).

Tsoi, K. K., Ng, S. S., Leung, M. C., & Sung, J. J. (2008). Cost-effectiveness analysis on screening for colorectal neoplasm and management of colorectal cancer in Asia. *Aliment Pharmacol Ther*, *28*(3), 353-363. https://doi.org/10.1111/j.1365-2036.2008.03726.x

Van Den Akker-Van Marle, M. E. (2002). Cost-Effectiveness of Cervical Cancer Screening: Comparison of Screening Policies. *CancerSpectrum Knowledge Environment*, *94*(3), 193-204. https://doi.org/10.1093/jnci/94.3.193

Van Der Meulen, M. P., Lansdorp-Vogelaar, I., Goede, S. L., Kuipers, E. J., Dekker, E., Stoker, J., & Van Ballegooijen, M. (2018). Colorectal Cancer: Cost-effectiveness of Colonoscopy versus CT Colonography Screening with Participation Rates and Costs. *Radiology*, *287*(3), 901-911. https://doi.org/10.1148/radiol.2017162359

van Hees, F., Habbema, J. D., Meester, R. G., Lansdorp-Vogelaar, I., van Ballegooijen, M., & Zauber, A. G. (2014). Should colorectal cancer screening be considered in elderly persons without previous screening? A cost-effectiveness analysis. *Ann Intern Med*, *160*(11), 750-759. https://doi.org/10.7326/m13-2263

van Hees, F., Zauber, A. G., Klabunde, C. N., Goede, S. L., Lansdorp-Vogelaar, I., & van Ballegooijen, M. (2014). The Appropriateness of More Intensive Colonoscopy Screening Than Recommended in Medicare Beneficiaries: A Modeling Study. *JAMA Internal Medicine*, *174*(10), 1568-1576. https://doi.org/10.1001/jamainternmed.2014.3889

Van Luijt, P. A., Rozemeijer, K., Naber, S. K., Heijnsdijk, E. A., Van Rosmalen, J., Van Ballegooijen, M., & De Koning, H. J. (2016). The role of pre-invasive disease in overdiagnosis: A microsimulation study comparing mass screening for breast cancer and cervical cancer. *Journal of Medical Screening*, *23*(4), 210-216. https://doi.org/10.1177/0969141316629505

van Rosmalen, J., de Kok, I. M., & van Ballegooijen, M. (2012). Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. *Bjog*, *119*(6), 699-709. https://doi.org/10.1111/j.1471-0528.2011.03228.x

van Rossum, L. G., van Rijn, A. F., Verbeek, A. L., van Oijen, M. G., Laheij, R. J., Fockens, P., Jansen, J. B., Adang, E. M., & Dekker, E. (2011). Colorectal cancer screening comparing no screening, immunochemical and guaiac fecal occult blood tests: a cost-effectiveness analysis. *Int J Cancer*, *128*(8), 1908-1917. https://doi.org/10.1002/ijc.25530

Vanness, D. J., Knudsen, A. B., Lansdorp-Vogelaar, I., Rutter, C. M., Gareen, I. F., Herman, B. A., Kuntz, K. M., Zauber, A. G., van Ballegooijen, M., Feuer, E. J., Chen, M. H., & Johnson, C. D. (2011). Comparative economic evaluation of data from the ACRIN National CT Colonography Trial with three cancer intervention and surveillance modeling network microsimulations. *Radiology*, *261*(2), 487-498. https://doi.org/10.1148/radiol.11102411

Vijan, S., Hwang, E. W., Hofer, T. P., & Hayward, R. A. (2001). Which colon cancer screening test? A comparison of costs, effectiveness, and compliance. *Am J Med*, *111*(8), 593-601. https://doi.org/10.1016/s0002-9343(01)00977-9

Vijan, S., Hwang, I., Inadomi, J., Wong, R. K., Choi, J. R., Napierkowski, J., Koff, J. M., & Pickhardt, P. J. (2007). The cost-effectiveness of CT colonography in screening for colorectal neoplasia. *Am J Gastroenterol*, *102*(2), 380-390. https://doi.org/10.1111/j.1572-0241.2006.00970.x

Vilaprinyo, E., Forné, C., Carles, M., Sala, M., Pla, R., Castells, X., Domingo, L., & Rue, M. (2014). Cost-effectiveness and harm-benefit analyses of risk-based screening strategies for breast cancer. *PLoS One*, *9*(2), e86858. https://doi.org/10.1371/journal.pone.0086858

Viscondi, J. Y. K., Faustino, C. G., Campolina, A. G., Itria, A., & Soárez, P. C. (2018). Simple but not simpler: a systematic review of Markov models for economic evaluation of cervical cancer screening. *Clinics (Sao Paulo)*, *73*, e385. https://doi.org/10.6061/clinics/2018/e385

Wagner, J. L., Herdman, R. C., & Wadhwa, S. (1991). Cost effectiveness of colorectal cancer screening in the elderly. *Ann Intern Med*, *115*(10), 807-817. https://doi.org/10.7326/0003-4819-115-10-807

Wan, L., Lou, W., Abner, E., & Kryscio, R. J. (2016). A comparison of time-homogeneous Markov chain and Markov process multi-state models. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *2*(3-4), 92-100. https://doi.org/10.1080/23737484.2017.1361366

Wang, Z. H., Gao, Q. Y., & Fang, J. Y. (2012). Repeat colonoscopy every 10 years or single colonoscopy for colorectal neoplasm screening in average-risk Chinese: a cost-effectiveness analysis. *Asian Pac J Cancer Prev*, *13*(5), 1761-1766. https://doi.org/10.7314/apjcp.2012.13.5.1761

Watts, R. D., & Li, I. W. (2019). Use of Checklists in Reviews of Health Economic Evaluations, 2010 to 2018. *Value Health*, *22*(3), 377-382. https://doi.org/10.1016/j.jval.2018.10.006

Weinstein, M. C. (1990). Principles of cost-effective resource allocation in health care organizations. *Int J Technol Assess Health Care*, *6*(1), 93-103. https://doi.org/10.1017/s0266462300008953

Weinstein, M. C. (2006). Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics*, *24*(11), 1043-1053. https://doi.org/10.2165/00019053-200624110-00002

Whyte, S., Chilcott, J., Cooper, K., Essat, M., Stevens, J., Wong, R., & Kalita, N. (2011). *Re-appraisal of the options for colorectal cancer screening. Report. Report for the NHS Bowel Cancer Screening Programme. National Health Service.*

Whyte, S., Chilcott, J., & Halloran, S. (2012). Reappraisal of the options for colorectal cancer screening in England. *Colorectal Dis*, *14*(9), e547-561. https://doi.org/10.1111/j.1463-1318.2012.03014.x

Whyte, S., Thomas, C., Chilcott, J., & Kearns, B. (2022). Optimizing the Design of a Repeated Fecal Immunochemical Test Bowel Cancer Screening Programme With a Limited Endoscopy Capacity From a Health Economic Perspective. *Value Health*, *25*(6), 954-964. https://doi.org/10.1016/j.jval.2021.10.002

Williams, C., Lewsey, J. D., Briggs, A. H., & Mackay, D. F. (2017). Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making*, *37*(4), 340-352. https://doi.org/10.1177/0272989x16651869

Wilschut, J. A., Habbema, J. D. F., Van Leerdam, M. E., Hol, L., Lansdorp-Vogelaar, I., Kuipers, E. J., & Van Ballegooijen, M. (2011). Fecal Occult Blood Testing When Colonoscopy Capacity is Limited. *JNCI: Journal of the National Cancer Institute*, *103*(23), 1741-1751. https://doi.org/10.1093/jnci/djr385

Wilschut, J. A., Hol, L., Dekker, E., Jansen, J. B., Van Leerdam, M. E., Lansdorp–Vogelaar, I., Kuipers, E. J., Habbema, J. D. F., & Van Ballegooijen, M. (2011). Cost-effectiveness Analysis of a Quantitative Immunochemical Test for Colorectal Cancer Screening. *Gastroenterology*, *141*(5), 1648-1655.e1641. https://doi.org/10.1053/j.gastro.2011.07.020

Wilson, J. M. G., & Jungner, G. (1968). *Principles and practice of screening for disease*. World Health Organization. https://apps.who.int/iris/handle/10665/37650

Wolff, H. B., Qendri, V., Kunst, N., Alarid-Escudero, F., & Coupé, V. M. H. (2022). Methods for Communicating the Impact of Parameter Uncertainty in a Multiple-Strategies Cost-Effectiveness Comparison. *Med Decis Making*, *42*(7), 956-968. https://doi.org/10.1177/0272989x221100112

Wong, C. K., Lam, C. L., Wan, Y. F., & Fong, D. Y. (2015). Cost-effectiveness simulation and analysis of colorectal cancer screening in Hong Kong Chinese population: comparison amongst colonoscopy, guaiac and immunologic fecal occult blood testing. *BMC Cancer*, *15*, 705. https://doi.org/10.1186/s12885-015-1730-y

World Health Organization, Baltussen, R. M. P. M., Adam, T., Tan-Torres Edejer, T., Hutubessy, R. C. W., Acharya, A., Evans, D. B., Murray, C. J. L., & Who, C. (2003). Making choices in health : WHO guide to cost-effectiveness analysis / edited by T. Tan-Torres Edejer ... [et al]. In. Geneva: World Health Organization.

Wu, G. H., Wang, Y. M., Yen, A. M., Wong, J. M., Lai, H. C., Warwick, J., & Chen, T. H. (2006). Cost-effectiveness analysis of colorectal cancer screening with stool DNA testing in intermediate-incidence countries. *BMC Cancer*, *6*, 136. https://doi.org/10.1186/1471-2407-6-136

Zhong, G. C., Sun, W. P., Wan, L., Hu, J. J., & Hao, F. B. (2020). Efficacy and cost-effectiveness of fecal immunochemical test versus colonoscopy in colorectal cancer screening: a systematic review and meta-analysis. *Gastrointest Endosc*, *91*(3), 684-697.e615. https://doi.org/10.1016/j.gie.2019.11.035

ZIN. (2016). *Guideline for economic evaluations in healthcare*. Zorginstituut Nederland. https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare

APPENDIX 1 A SYSTEMATIC REVIEW OF COST-EFFECTIVENESS ANALYSES OF COLORECTAL CANCER SCREENING IN EUROPE: HAVE STUDIES INCLUDED OPTIMAL SCREENING INTENSITIES?

Abstract

Objective*:* To assess the range of strategies analysed in European cost-effectiveness analyses (CEAs) of colorectal cancer screening (CRC) screening with respect to the screening intervals, age ranges and test cut-offs used to defined positivity to examine how this might influence what strategies are found to be optimal, and consider the implications within the context of current screening policies with a focus on the screening interval.

Methods: We searched PubMed, Web of Science, and Scopus for peer-reviewed, model-based CEAs of CRC screening. We included studies on average-risk European populations using the guaiac fecal occult blood test (gFOBT) or fecal immunochemical test (FIT). We adapted Drummond's 10-point checklist to appraise study quality.

Results: We included 39 studies that met the inclusion criteria. Biennial screening was the most frequently used interval which was analysed in 37 studies. Annual screening was assessed in 13 studies, all of which found it optimally cost-effective. Despite this, 25 of 26 European stool-based programmes use biennial screening. Many CEAs did not vary the age range, but the 14 that did generally found broader ranges optimal. Only 11 studies considered alternative FIT cut-offs, 9 of which found lower cut-offs superior. Conflicts between current policy and CEA evidence are less clear regarding age ranges and cut-offs.

Conclusions: The existing CEA evidence indicates that the widely adopted biennial frequency of stool-based testing in Europe is sub-optimal. It is likely that many more lives could be saved throughout Europe if programmes move to more intensive annual screening.

Highlights

What is already known about the topic?

The relevance of multiple possible alternative intervention intensities has implications for the identification of optimally cost-effective strategies has long been recognised within cost-effectiveness analysis (CEA). Thus, when estimating the cost-effectiveness of CRC screening analysts should take care to include sufficient variation in screening age range, screening intervals and test positivity cut-offs.

What does the paper add to the existing knowledge?

Few European CEAs of CRC stool-based screening include wide ranges of strategies in their analyses; the minority of CEAs that assessed annual screening all found it to be optimally cost-effective. In contrast, majority of the European screening programmes use biennial screening interval. This mismatch between the policy and evidence likely indicate the European screening programmes are sub-optimal. Potentially many more lives could be saved throughout Europe if programmes move to more intensive annual screening.

What insights does the paper provide for informing healthcare-related decision making?

Modellers need to better understand and adhere to existing guidance on strategy selection in CEAs of CRC screening. They should include a wide range of strategies beyond those examined in trials or recommended within guidelines to correctly identify optimal strategies and best inform decision-makers. In the specific case of CRC screening, annual screening should be considered in CEAs.

Introduction

Colorectal cancer (CRC) is a major public health problem in Europe. It is the most common cancer in terms of incidence and second in terms of mortality with 520,000 new cases and 245,000 deaths annually (1) with an estimated economic burden of ~ €19 billion (2). CRC screening can reduce mortality by detecting disease at an earlier, more treatable stage (3). Stool-based screening modalities are non-invasive tests such as the guaiac fecal occult blood test (gFOBT), fecal immunochemical test (FIT) and fecal DNA test. Imaging tests are invasive and employ either endoscopic approaches such as colonoscopy, sigmoidoscopy or non-endoscopic techniques such as computed tomography colonoscopy, magnetic resonance colonoscopy or colon capsule endoscopy (4).

In 2003, the Council of the European Union (EU) recommended population-based CRC screening between ages 50–74 years using gFOBT (5). Many European countries have since initiated or expanded population-based CRC screening programmes (4). Programmes have typically used gFOBT or FIT as the primary screening test and then employ colonoscopy to triage primary screen positives. FIT has superior test sensitivity to gFOBT and the advantage of the ability to quantitatively adjust the sensitivity and specificity trade-off by varying the FIT cut-off threshold for test positivity (6). Thus, many programmes have switched from gFOBT to FIT (7).

Like many other screening programmes, the intensity of CRC screening can be varied by adjusting the screening start age, stop age and screening interval. CRC screening is cost-effective when offered at an appropriate intensity (6). The relevance of multiple possible alternative intervention intensities has implications for the identification of optimally cost-effective strategies that have been long-recognised within cost-effectiveness analysis (CEA) (8). Problems can arise if CEAs include insufficient strategies. A study may conclude a given strategy is cost-effective but may have failed to consider other relevant strategies, including those that are potentially more effective and cost-effective (6). Thus, when interpreting cost-effectiveness of estimates of CRC screening it is important to consider whether analyses included sufficient variation in screening age range, screening intervals and FIT cut-offs to provide the most useful policy guidance.

The primary objective of our review is to assess the range of screening intensities considered in European CEAs regarding the screening age ranges, screening intervals and FIT cut-offs and consider what implications this had for what was found to be optimally cost-effective. In particular, we aim to assess how many CEAs included annual screening and to examine if it is the optimal interval in those studies that have included it. Secondly, we aim to assess the intensity of current European CRC screening policies in the context of the current cost-effectiveness evidence regarding optimal screening strategies. We are not aware of any prior such examination of the

CEA evidence of CRC screening regarding the optimally cost-effective choice and European screening policies.

Methods

Our review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (9). The protocol is registered on PROSPERO (CRD42021283988), an international register of prospective systematic reviews.

Search strategy

We conducted a systematic search of the PubMed, Web of Science, and Scopus databases on 9th September 2022. The search strings are provided in the supplementary material 1. We also manually reviewed the reference lists of 7 relevant systematic reviews on CEAs of CRC screening (10-16). Characteristics of the European screening programmes were identified from recent academic publications and by contacting experts.

Eligibility criteria

We included peer-reviewed CEAs of CRC screening in average-risk populations in the 27 EU member states and the UK (henceforth collectively referred to as Europe). The inclusion criteria were; (i) model-based CEA simulations of CRC screening using stool-based tests; (ii) conducted in average-risk populations; (iii) set in Europe. The exclusion

criteria were studies: (i) of stool-based tests other than FIT or gFOBT such as stool DNA testing; (ii) that did not report quality-adjusted life-years (QALYs) or life-years gained (LYG) or equivalents such as life-years saved; (iii) that did not report costs; (iv) not of a screening intervention; (v) not written in English; and (vii) reviews, conference abstracts, editorials and notes.

Study selection

The titles and abstracts of studies from the database searches and other sources were recorded in Microsoft Excel (version 2108) and duplicates were removed. Records were independently screened by two reviewers (RP and YL). A third reviewer (JOM) was consulted in case of disagreement. Full text of studies that met the inclusion criteria were exported to the Mendeley reference management software (version 2.57) and examined in full. RP, YL, EM and JOM assessed the studies based on the quality assessment checklist described below.

Data extraction

Two reviewers (RP and YL) individually extracted the study publication year, authors, country setting, details of the simulated strategies such as screening start and stop ages, intervals between screens, test modality, costs, estimated QALYs or LYG, reported incremental cost-effectiveness ratio (ICER) and cost-effectiveness threshold (if stated). As not all studies report a threshold, we used the threshold reported in other studies from the same setting as a proxy. Where there were multiple studies from the same setting that reported different thresholds we used the highest as the proxy threshold. We interpret the optimally cost-effective strategy in the conventional way as being the most

effective strategy with an ICER within the cost-effectiveness threshold. A third reviewer (JOM) checked a random 20% of the extracts for verification.

Assessment of quality of studies

A quality assessment checklist was developed based on Drummond's 10-item checklist (17), with adjustments corresponding to our study aims. The adjusted checklist is attached in supplementary material 2. The adjusted Drummond checklist was used to assess the quality of the reviewed studies. We use the same 10 topic areas as the original Drummond checklist but omitted 9 of the 33 specific sub-questions as they did not apply to the studies considered here (17). Each study was graded on the 10 topic areas, with a 1 and 0.5 being awarded where the study fully or partially satisfied the checklist questions respectively and 0 where the questions were not satisfied. Grades were totalled to generate a summary score.

Results

Our search identified 984 studies yielding 430 unique studies after removing duplicates. Following title and abstract screening 382 studies were excluded. We assessed the full texts of the remaining 56 studies and 7 additional studies identified from manual reference checking. A total of 39 studies were ultimately included (6, 18-56). A PRISMA diagram is presented in Figure 1.

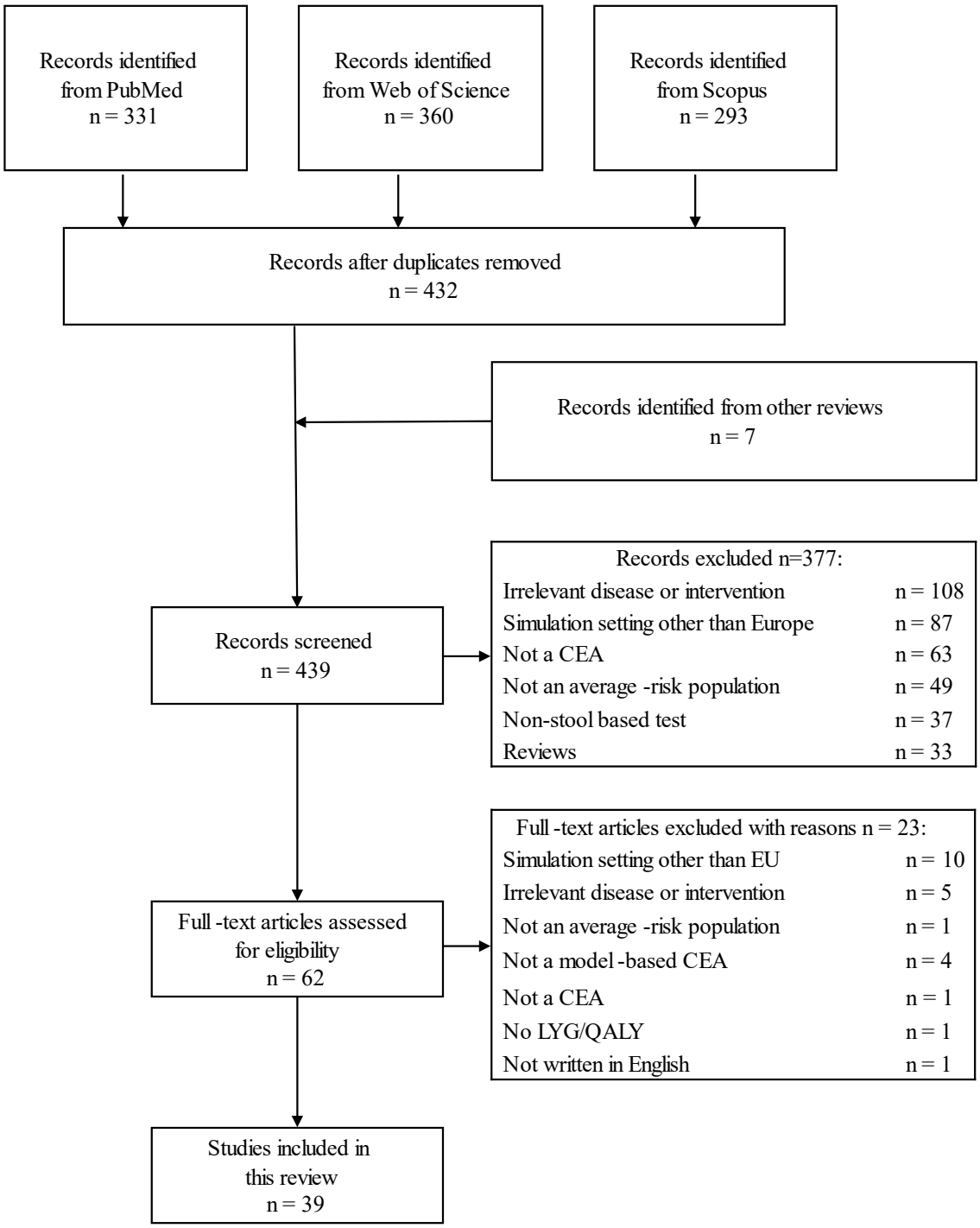


Figure 1 – The PRISMA flow diagram of the literature search and selection process.

PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Table 1: Characteristics of CEAs conducted in Europe

First Author Reference, n= 39	Simulation setting	Stool-based tests included	Screening interval (years)	FIT cut-off (µg Hb/g)	Screening start and stop age (interval)	Number of stool-based strategies	CEA Cost-effectiveness threshold reported (/QALY)	Optimally cost-effective strategy (Test name, interval, FIT cut-off, age range)
Aeria 2019	Portugal	FIT	2	20	50-74	2	€39,760	FIT, biennial, 20, 50-74
Aronsson 2017	Sweden	FIT	2, twice in lifetime	10	60-80	2	€10,000	Colonoscopy, 10 years, 60-80
Arrospide 2018	Spain	FIT	2	20	50-69	1	*£30,000	FIT, biennial, 20, 50-69
Babela 2021	Slovakia	FIT	1,2	20	50-75		$50,000	FIT, annual, 20, 50-75
Barre 2020	France	FIT	2	30	50-74	3	€40,000	FIT, biennial, 30, 50-74
Berchi 2004	France	FIT, FOBT	2	Not given	50-74	4	*€50,000	FIT, biennial, NG, 50-74
Chauvin 2012	France	FIT, FOBT	2	Not given	50-80	2	*€50,000	FIT, biennial
Coretti 2020	Italy	FIT	2	20	50-70	1	€30,000	FIT, biennial, NG, 50-70
Currais 2021	Portugal	FIT	2	20	45-	1	€39,760	None
Goede 2013	The Netherlands	FIT	1, 1.5, 2 ,3	10,15,20,30,40	45/ 50/ 55/ 60 - 70/75/80	960	*€50,000	FIT, annual, 10, 45-80
Greuter 2016	The Netherlands	FIT	2	15	55-75	1	*€50,000	FIT, biennial, 15, 55-75
Gyrd Hansen 1998	Denmark	FIT, FOBT	2,1	Not given	50/55-74	5	*DKK 88000	FOBT, annual, NA, 50-74
Gyrd Hansen 1998b	Denmark	FOBT	3,2,1.5,1	NA	50/55/60/65/70-54/59/64/69/74	60	*DKK 88000	FOBT, annual, NA, 50-74
Hassan 2011	France	FIT, FOBT	1, 2	Not given	50-75	4	$50,000	FIT, annual, NG, 50-75
Heinävaara 2022	Finland	FIT	1,2,3	10,25,40,55,70	50/55/60/65-69/71/79	362	€10,000	FIT, annual, 25,50-79 (m),10, 55-69 (w)
Heresbach 2010	France	FIT, FOBT	2	Not given	50-74	2	€10,000 & 20,000	FIT, biennial, NG, 50-74
Idigoras 2018	Spain	FIT	2	20	50-69	1	*£30,000	FIT, biennial,20, 50-69
Jahn 2019	Austria	FIT, FOBT	1	Not given	50-75	2	€15,000	FIT, annual, NG, 50-75
Ladabaum 2014	Germany	FIT, FOBT	1, 2	Not given	50-54, once at 50 and 60	5	€25,000 & 50,000	FIT, annual (50-54) followed by biennial up to 75
Lansdorp-Vogelaar 2018	The Netherlands	FIT	1,1.5,2,3,5,7,10	10	40/50/55/60-70/75/80	84	€50,000	FIT, annual, 10, 45-80
Lee 2010	The UK	FOBT	2	NA	60-69	1	£30 000	CTC, 10 yearly
Lejeune 2004	France	FOBT	2	NA	50-74	2	$20,000	FOBT, biennial, NA, 50-74
Lejeune 2010	France	FIT, FOBT	2	Not given	50-74	2	£20,000	FIT, biennial, NG, 50-74
Lejeune 2014	France	FIT, FOBT	2	4,30,35,35.2,40,41, 46.8,50,58.6,60, 70.4	50-74	15	£20,000–30,000	FIT (2 samples), biennial, 35.2, 50-74
Macafee 2008	The UK	FOBT	2	NA	60-69	1	*£20,000	FOBT, biennial, NA, 60-69

First Author Reference, n= 39	Simulation setting	Stool-based tests included	Screening interval (years)	FIT cut-off (µg Hb/g)	Screening start and stop age (interval)	Number of stool-based strategies	CEA Cost-effectiveness threshold reported (/QALY)	Optimally cost-effective strategy (Test name, interval, FIT cut-off, age range)
McFerran 2022	Ireland	FIT	1,2,3,4,5	10,15,20,30,40	45/50/55/60/65/70-70/75/80	525	€20,000	FIT, annual, 10, 50-80
Murphy 2017	The UK	FIT, FOBT	2	20, 40, 100, 150, 180	60-74	6	£20,000	FIT, biennial, 20, 60-74
Pil 2016	Belgium	FIT	2	15	56-74	4	€35,000	FIT, biennial, NG, 55-74
Senore 2019	Italy	FIT	2	20	58-69	1	$50,000	FS at age 58 +biennial FIT for all non-attendees to FS
Sharp 2012	Ireland	FIT, FOBT	2	20	55/65-64/74	6	€45,000	FIT, biennial, 20, 55-74
Sobhani 2011	France	FIT, FOBT	2	10,15,20,30	50-62	2	£30,000	FIT (3 sample), biennial, 10, 50-62
Tappenden 2007	The UK	FOBT	2	NA	50/60/61-69/70/74	5	*£20,000	FS age 60, FOBT, biennial, NA, 61-70
Thomas 2021	The UK	FIT	2	20,80,120	50/54/58/60-74	25	£20 000	FIT, biennial, 120, 56(m)/60(w)-74
Van der Meulen 2017	The Netherlands	FIT	1,1.5,2,3	10,15,20,30,40	40/45/50/55/60/65-70/75/80/85	480	€20,000	FIT, annual, 10, 45-80
Van Rossum 2011	The Netherlands	FIT, FOBT	Once in lifetime	20	50-75	2	€25,000-100,000	FIT, once in a lifetime, 20, 50-75
Whyte 2012	The UK	FIT, FOBT	2	20	56/60/66-69/74	9	£20,000	FS age 55, FIT, biennial, 20, 56–74
Whyte 2021	The UK	FIT	1,2,3,4,5,6	4-36	50/51/52/53/54/55/56/57/58/59/60-74	60,221	£20,000	FIT, annual, 27-32.8, 50/51-71/74
Wilschut 2011	The Netherlands	FIT	1,1.5,2	10,15,20,30,40	45/50/55/60-70/75/80	48	€20000	FIT, annual, 10, 45-80
Wilschut 2011b	The Netherlands	FIT, FOBT	1,1.5,2	10,15,20,30,40	45/50/55/60-70/75/80	48	*€50,000	FIT, annual, 10, 45-80

FIT- Fecal Immunochemical Test, FOBT- Fecal Occult Blood Test, FS-Flexible Sigmoidoscopy, CTC- Computed Tomography Colonoscopy QALY- Quality Adjusted Life Years, *- CEA threshold cited in other literature from the same country, NA-Not Applicable, NG- Not Given, µg Hb/g- Microgram haemoglobin per gram, DKK-Danish krone

CEAs of stool-based tests were simulated in 14 different countries; France, the Netherlands, and the UK were the most frequent source countries with 9, 7 and 7 studies respectively. Slovakia is the only country within the sample from Central and Eastern Europe as defined by the Organisation of Economic Cooperation and Development (57). Most studies (n=33, 85%) were conducted after 2010. Of the 39 studies, 37 studies (95%) were published after the European Council's 2003 recommendations.

Of the 39 CEAs, 22 (56%) exclusively analysed stool-based tests while 17 (44%) also assessed image-based primary screening modalities. Of those that exclusively simulated stool-based tests, 12 assessed FIT only, 3 assessed gFOBT only and 6 assessed both FIT and gFOBT, and 1 assessed a biomarker test. Nearly all studies assessed strategies in which recipients receive the same modality over the entire screening programme. One exception was set in Germany, in which younger screenees start with stool-based testing but later switch to image-based testing (35).

While most studies included some variation of screening strategies regarding the screening age range, interval or FIT cut-off, the level of variation was limited. There were 26 (67%) studies that only simulated five strategies or fewer.

Figure 2 illustrates the screening intervals considered by the reviewed studies. Regarding the screening interval, 14 studies (36%) examined more than one screening interval. Biennial screening was the most frequently simulated interval which was

assessed in 37 (95%) studies of which 23 (59%) considered no other interval, assessing only biennial screening. Whilst annual screening was assessed in only 13 (33%) studies, the majority of these (n=12) also considered other intervals.

Figure 2 uses a white diamond to highlight the intervals found to be optimally cost-effective in those studies which considered multiple intervals. All 13 studies that assessed annual screening intervals found it to be optimally cost-effective. That is, annual screening was the most effective strategy with an ICER within the CEA threshold in all cases. Naturally, that annual screening was always found to be optimally cost-effective does not imply that all strategies featuring annual screening are cost-effective. Four studies that varied the screening age range and FIT cut-offs found some strategies featuring annual screening with ICERs above the threshold (6, 27, 30, 58).

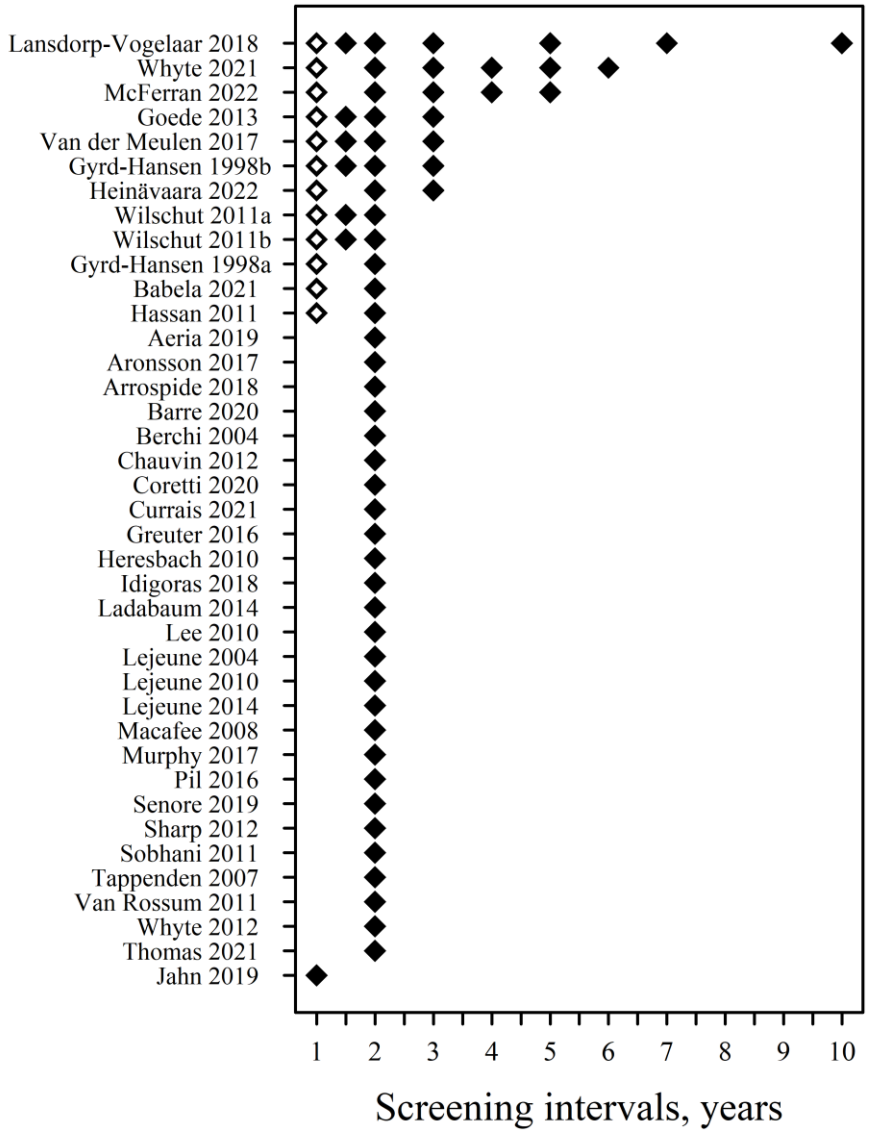


Figure 2 Screening interval analysed in the studies

Regarding the age ranges simulated, there was wide variation in the ranges examined by the studies as depicted in Figure 3. Of the 39 studies, 27 (69%) did not consider alternative age ranges. Ages 50-74 was the most commonly analysed range, which was assessed in 11 (28%) studies followed by 50-75 used in 9 (23%). Only 7 (18%) included start ages below 50 and 12 (31%) analysed stop ages above 75. Figure 3 uses white diamonds to indicate the start and stop age of those age ranges found to be

optimal within those studies that did consider alternative age ranges. All such studies found starting and stopping between 45-50 and 74-80 respectively to be optimal.

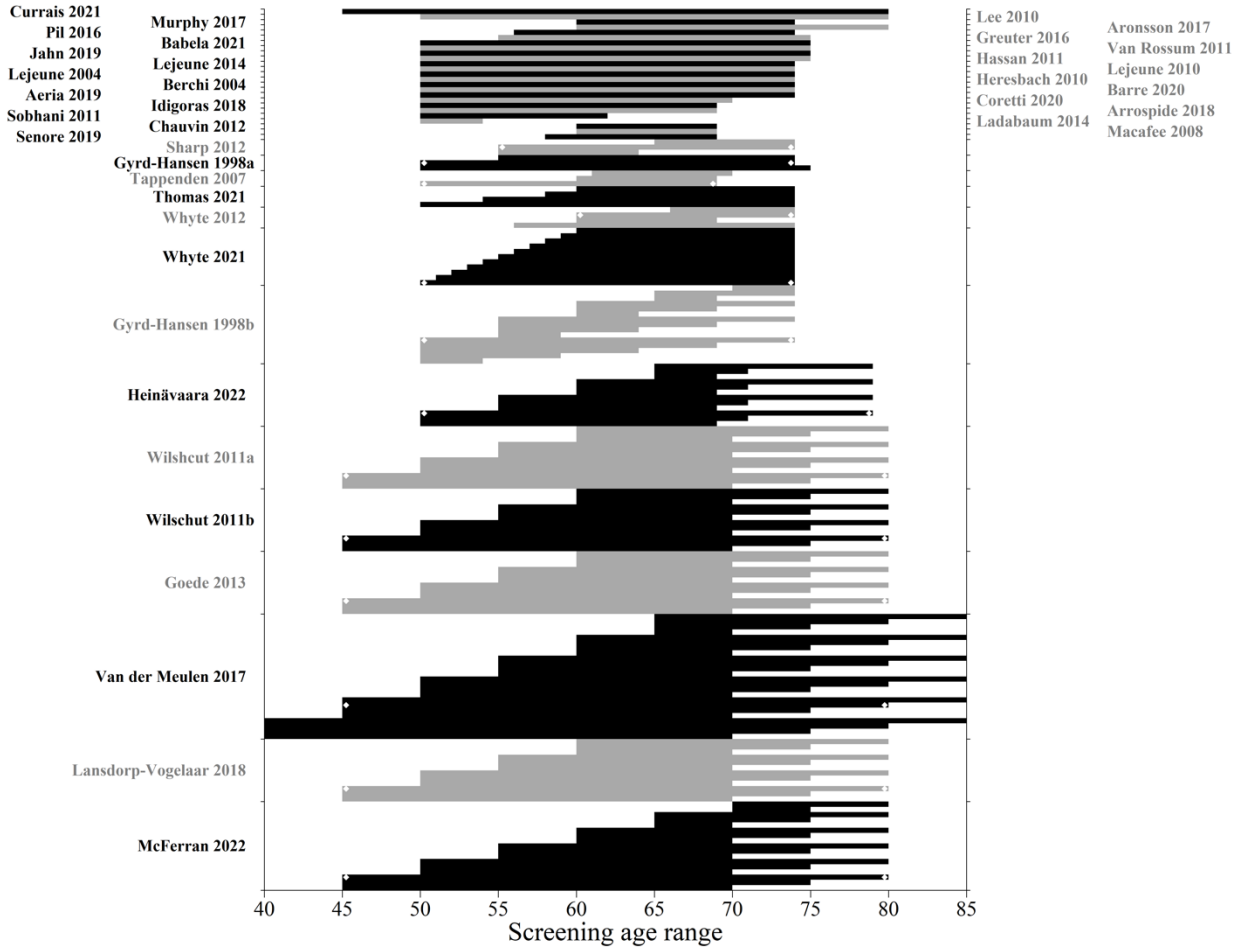


Figure 3 Age range analysed in the studies

Note: White diamonds indicate the optimally cost-effective strategies in all Figures

The FIT cut-off was reported in 26 of the 33 studies (79%) examining FIT-based strategies. Figure 4 presents the FIT cut-offs considered. The most common cut-offs used were 20 and 10 µg Hb/g of faeces with 18 and 10 studies respectively (55% and 30%). Multiple cut-offs were examined in 11 (28%) studies. Of these, 9 found lower cut-offs dominated, the exceptions being studies from the UK and France (39, 51).

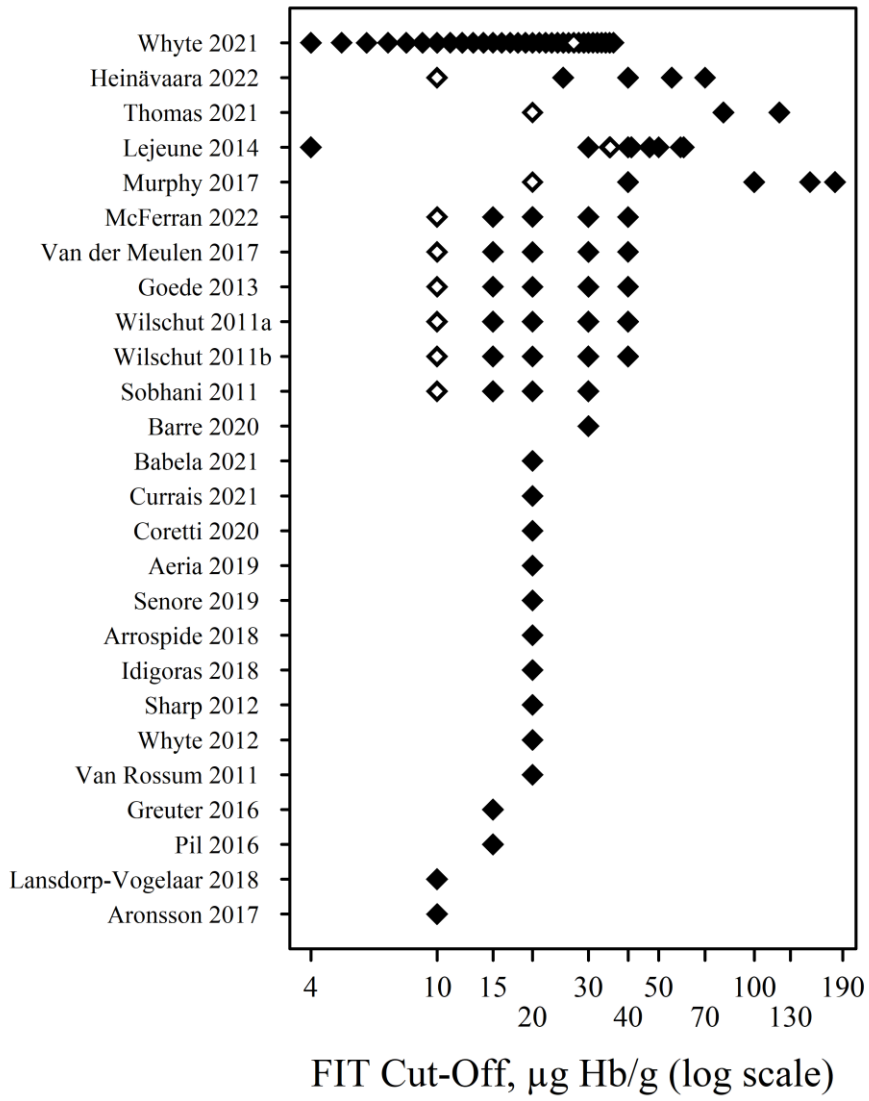


Figure 4 FIT cut-off analysed in the studies

Table 2: European CRC screening programmes characteristics

Country	Screening programme type	Start year	Primary test	Screening interval (years)	FIT cut off (µg Hb/g)	Screening start & stop age (years)	Triage test	References
Austria Burgenland	Organised	2003	FIT	1	10	40-80	Colonoscopy	(57)
Austria Vorarlberg	Organised	2007	Colonoscopy	10	NA	>50years		(57)
Belgium Flanders	Organised	2013	FIT	2	15	50-74	Colonoscopy	(58)
Belgium Wallonia	Organised	2009	FIT	2	15	50-74	Colonoscopy	(59)
Bulgaria	Opportunistic	NA	NA	NA	NA	NA	NA	
Croatia	Organised	2007	FOBT	2	NA	50-74	Colonoscopy	(60)
Cyprus (pilot/planned)	Organised	2013	FIT	2	Not given	50-69		
Czech Republic	Organised	2000	FIT TC	1 or 2*	15	50+	Colonoscopy	
Denmark	Organised	2014	FIT	2	20	50-74	Colonoscopy	(61)
Estonia (pilot/planned)	Organised	2016	FIT	2	Not given	60-69		(62)
Finland	Organised	2009	FIT	2	25 (women), 70 (men)	60-69	Colonoscopy	(63)
France	Organised	2009	FIT	2	30 µg/ml	50-74	Colonoscopy	(21)
Germany	Organised	1971	FIT/TC	1 or 2*	4,6,8,10,15,17,25 µg/ml	50 (men), 55 (women)	Colonoscopy	(64)
Greece	Opportunistic	NA	NA	NA	NA	NA	NA	
Hungary	Organised	2018	FIT	2	20	50-70	Colonoscopy	(65)
Ireland	Organised	2012	FIT	2	45	60-69	Colonoscopy	(43)
Italy	Organised	1982	FIT/FS	2	20	50-69	Colonoscopy	(23)
Latvia	Opportunistic	NA	NA	NA	NA	NA	NA	
Lithuania	Organised	2014	FIT	2	6/40 µg Hb/g or 20 µg Hb/ml	50-74	Colonoscopy	(66)
Luxembourg	Organised	2016	FIT TC	2	Not given	55-74	Colonoscopy	(62)
Malta	Organised	2013	FIT	2	16-20	55-66	Colonoscopy	(62)
The Netherlands	Organised	2014	FIT	2	47	55-75	Colonoscopy	(67)
Poland	Organised	2012	Colonoscopy	Once in a lifetime	n/a	55-64		(62)
Portugal	Organised	2009	FIT	2	20	50-70		(25, 62)
Romania	NA	NA	FIT	2	NG	NA	NA	
Slovakia	Organised	2019	NA	NA	NA	NA	NA	(20)
Slovenia	Organised	2009	FIT	2	20	50-74	Colonoscopy	(68)
Spain Basque	Organised	2008	FIT	2	20	60-69	Colonoscopy	(32)
Sweden (Stockholm/Gotland County)	Organised	2008	FIT	2	40 (women), 80 (men)	60-69	Colonoscopy	(69)
UK-England**	Organised	2006	FIT	2	120	60-74	Colonoscopy	(70)
UK-Wales**	Organised	2007	FIT	2	150	55-74	Colonoscopy	(60)
UK-Scotland**	Organised	2008	FIT	2	80	50-74	Colonoscopy	(71)
UK-Northern Ireland**	Organised	2010	FIT	2	150	60-74	Colonoscopy	(72)

*50-54-annual, 55- biennial, FIT- Fecal Immunochemical Test, FOBT- Fecal Occult Blood Test, FS-Flexible Sigmoidoscopy, TC- Total Colonoscopy, µg Hb/g- Microgram haemoglobin per gram, µg Hb/ml- Microgram haemoglobin per mililitre ** These UK nations previously conducted FOBT based screening programmes, and had phased introduction of FIT from 2016 onwards, Wales approved lowered start age to 55 years in October 2022.

Endoscopy capacity constraints

The potential for endoscopy capacity to constrain the feasibility of screening strategies has been recognised by several studies. Colonoscopy capacity is mentioned as an important consideration by 20 studies (48%) (6, 18-20, 24, 27, 33, 36, 39, 41, 44, 47-52, 55, 58). Of these, 4 studies actively considered strategies under capacity constraints within their analyses (6, 49-51). Those studies found annual screening was not optimally cost-effective in constrained analyses but was optimal in analyses without capacity constraints.

European CRC screening provision

Population-based screening is implemented in 22 of 28 European countries at national and regional levels. Cyprus and Estonia have pilot programmes. The countries without population-based programmes are Romania, Bulgaria, Greece, Latvia.

Most countries use stool-based primary testing except Poland and the Austrian region of Vorarlberg which both use primary colonoscopy. All countries using stool-based testing employ FIT except Croatia, where gFOBT test is used. Colonoscopy is used as triage in all countries with stool-based primary testing.

Among the 26 countries with stool-based screening including pilot programmes, 23 screen biennially. The three exceptions are Austria region of Burgenland, which screens annually, and Germany and the Czech Republic, which reduce screening frequency with age. Germany screens annually between ages 50-54 and then switches to biennial

screening. Similarly, the Czech programme screens annually between 50-54 and participants are then offered biennial FIT or 10-yearly colonoscopy.

Screening start and stop ages also vary. Vorarlberg (in Austria) starts at 40; 14 countries start at 50 and 7 programmes start at 60. Estonia, Finland, Ireland, Spain and Sweden had the narrowest population age range covered of 60-69 years. Sex-specific start ages apply in Germany at 50 and 55 for men and women respectively.

There is considerable variation in the FIT cut-off used across the EU. All countries use defined FIT thresholds except Malta, which uses a range of 16-20 µg Hb/g. Germany and Lithuania use different cut-offs depending on the test manufacturer (59, 60). Programmes using higher cut-offs are Ireland and Netherlands with 45 and 47 µg Hb/g respectively and the constituent countries of the UK, where 80-150 µg Hb/g is used. Programmes using lower cut-offs are Germany and Lithuania. The most common cut-off is 20 µg Hb/g which is used in 6 programmes. Sex-specific cut-offs are used in Finland and Sweden.

Comparison of policy to CEA evidence

The most apparent overall observation when comparing the CEA estimates with European screening policies is that although studies clearly indicate annual screening is optimally cost-effective when using FIT many screening programmes use biennial screening in practice, which is both less effective and less cost-effective. Studies indicating the superiority of annual screening have been published in the setting of the Netherlands, France, Denmark and Austria. In practice, while Burgenland (Austria)

uses annual screening, it has done so since 2002, long before the Austrian CEA was published (34). In the Netherlands, several Dutch CEAs were conducted prior to the implementation of organised screening in 2014 (27, 48, 49, 58). All of those that assessed annual screening found it cost-effective, yet the programme has employed biennial screening since its introduction, likely reflecting colonoscopy capacity constraints. Similarly, in France, while 8 of 9 CEAs only considered biennial screening (22, 23, 25, 32, 37-39, 45), one study did include annual screening and found it to be cost-effective (31). Despite this, France has used biennial screening since 2002.

Regarding screening age ranges, CEA estimates suggest start ages of 50 or lower to be optimal. In practice, 19 of 33 programmes feature screening start ages of 50 or below. Six programmes use a screening start age of 60 despite no CEA evidence which indicates it to be optimal. Ten programmes are using a stop age of 70 or below contrary to the CEA evidence of a stop age between 74 to 80 being optimal.

Regarding the FIT cut-off, while most simulation estimates indicate 10 µg Hb/g to be optimal only Burgenland uses such a low threshold. Germany and Lithuania both feature cut-offs as low as 4 µg Hb/g, but not on a programme-wide basis as there is variation in the cut-offs used within these systems. All Dutch CEA estimates report 20 µg Hb/g or lower to be optimal but the Dutch programme uses a higher cut-off of 47 µg Hb/g.

Quality appraisal of the studies included in the review

The quality appraisal checklist findings are reported in the supplementary material 2. Of the 39 studies, 2 achieved positive answers to all 10 checklist questions, and the remaining studies recorded 7.5/10 or higher overall. Principal areas of weakness were the comprehensive description of the competing alternatives and presentation and the description of the study results. Studies generally performed well on the remaining items.

Discussion

Our review assessed the screening intensities analysed by European CEAs of stool-based CRC screening. We found that nearly all CEAs examined biennial screening while only a minority considered annual screening. Every study that assessed both intervals found annual screening to be both more effective and cost-effective. Despite this, most European screening programmes currently use biennial screening. This indicates the widely adopted biennial screening interval in Europe is likely of sub-optimal intensity. That many studies omitted what appears to be the optimal screening interval from their analyses may have led policymakers across Europe to adopt an insufficiently intensive screening frequency.

Why many CEA analysts have simply chosen to simulate biennial screening rather than consider broader options is unclear. The 2003 European Council recommendation to adopt colorectal screening did not specify a particular interval (5). The three European randomised controlled trials (RCTs) of gFOBT published at that time had all examined a biennial interval (61-63). It is possible that health economists simply reflected the trialled intervals in their analyses. Despite this, a US trial completed in 1992 included

both annual and biennial screening and found the former more efficacious (64). Indeed, North American RCTs and CEAs assessing stool-based testing have mostly analysed annual screening (14, 15). Furthermore, annual testing is typically recommended by North American guidelines when stool-based testing is employed (65).

Existing CEA methods guidance strongly encourages modellers to simulate as broad a range of strategies as possible in order to correctly identify optimally cost-effective strategies (8). Indeed, it has long been recognised that considering a broad range of strategies is of particular relevance to CEAs of cancer screening (8). Furthermore, one of the motivating rationales of simulation modelling is that it permits the comparison of additional strategies not assessed in trials (66). Moreover, it is noteworthy that the earliest study within our review by Gyrd-Hansen from 1998 examined a range of intervals including annual screening (29, 30). In summary, the omission of annual screening by many studies is, despite relevant trial evidence, well-established methods guidelines and one early simulation study indicating annual screening is relevant for policy consideration.

Another possible explanation of why many modellers restricted their analysis to biennial screening could be tacit acceptance of colonoscopy constraints, not explicitly explored in their analyses. The Netherlands, UK, Finland and Ireland have faced colonoscopy capacity constraints and have been transparent in modifying aspects of screening provisions to operate within available capacity (27, 44, 51, 67). It is possible that analysts only considered biennial intervals as they anticipated there would be insufficient endoscopy capacity for annual screening. We believe, however, it is good

practice to simulate a broad range of strategies to correctly estimate the optimal strategy, with and without constraints. Such analyses have been published (6, 51), allowing modellers to demonstrate the shortcomings imposed by constraints and ensure policymakers are aware of the benefits of alleviating them.

The primary focus of this review has been the optimal interval, regarding which, it is quite clear that annual screening seems preferable. Analogous findings also apply regarding the screening age range and FIT cut-off, though the conclusions may be less definitive. CEAs that varied the screening age range generally find that broader age ranges are preferable. Similarly, many but not all studies that vary the FIT cut-off found lower cut-offs preferable. Accordingly, the conclusions regarding the need for strategy variation apply to age range and cut-off parameters also. The feasibility of optimising screening depends, in part, on available colonoscopy capacity. Moving from biennial to annual screening will increase colonoscopy demand, as will broadening the screening age range and reducing FIT cut-offs (27).

Our review shows that the simulation of restricted sets of policy alternatives is not merely a theoretical consideration but has meaningful policy significance. The omission of annual strategies from CEAs arguably could be interpreted as costing lives. That many studies have assessed an insufficient range of alternatives has resulted in incomplete cost-effectiveness evidence that may have misled decision-makers leading to inadvertently adopting sup-optimal policies, and failing to highlight how long-term capacity expansion planning should be prioritised in healthcare planning. Many thousands of lives might be saved across Europe if we could apply annual screening

and it is important that the issue of optimal policies is considered by modellers and policymakers.

While our review indicates that the likely optimal interval for FIT screening is annual is emerging evidence regarding risk-stratified screening based on age, sex, prior screening history, lifestyle and/or genetic information (68). Thus, the policies that provide optimally cost-effective CRC prevention might not necessarily require universal annual FIT screening and associated colonoscopy capacity. It seems likely that further research will address the potential for such tailoring and will likely support new policy recommendations without necessarily requiring large increases in colonoscopy capacity. The relevance of considering a breadth of strategies identified by our review will apply equally to research considering stratified policies.

An important caveat to our results is that the sample of studies from which there were CEAs to examine are drawn from Western Europe and thus predominately from wealthier European states. Moreover, the 13 studies within our sample that considered annual screening were from 7 wealthy European countries (Austria, Denmark, France, Finland, Ireland and the Netherlands and the UK). Furthermore, all five Dutch studies employed the same model. It is open to question whether annual screening would remain cost-effective when assessed among less wealthy countries, including those in Central and Eastern Europe.

Our review differs from previous systematic reviews of CEAs on CRC screening as those studies did not focus on the methodological consideration of the adequacy of the range of strategies. We focused on a specific question of the CEA evidence regarding optimal intervals for stool-based testing in the context of European screening policies. Thus, our review is limited compared to prior reviews in terms of the screening modalities examined and the context of simulations (10-16).

Our findings are based on reviewing the outcomes of modelling studies. We appreciate the uncertainty of simulation analyses. There may be scope for further RCTs to reduce uncertainties regarding the benefits of intensified screening, especially as there has been only one RCT comparing the biennial and annual intervals (64). The same applies regarding extending screening ages and lowering FIT cut-offs.

Limitations

Our study has several limitations. First, not all screening programmes clearly document screening strategy employed, especially the FIT cut-off used. Accordingly, we cannot be definitive about screening practices in all programmes. Second, we did not examine the model types or assumptions used in studies which might have played a role in the optimal strategy recommendations. Despite possible differences all relevant studies consistently found annual screening to be more effective and cost-effective than biennial screening. Third, we limited our analysis to stool-based based tests and did not consider image-based screening modalities or a combination of stool and image-based modalities.

Conclusion

The existing CEA evidence indicates that the widely adopted biennial stool-based testing in Europe is likely sub-optimal. Every CEA that examined annual screening found it to be optimally cost-effective. It is the responsibility of modellers to include a wide range of strategies beyond those examined in trials or recommended within guidelines to correctly identify optimal strategies and best inform decision-makers. CRC screening strategies, especially the intervals in FIT based testing should be reconsidered. It is likely that many more lives could be saved throughout Europe if programmes move to more intensive annual screening. The feasibility of providing such frequent screening depends however on the colonoscopy capacity within each health system.

References

1. Cardoso R, Guo F, Heisser T, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol*. 2021; 22: 1002-13.

2. Henderson RH, French D, Maughan T, et al. The economic burden of colorectal cancer across Europe: a population-based cost-of-illness study. *Lancet Gastroenterol Hepatol*. 2021; 6: 709-22.

3. Jodal HC, Helsingen LM, Anderson JC, et al. Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. *BMJ open*. 2019; 9: e032773.

4. Schreuders EH, Ruco A, Rabeneck L, et al. Colorectal cancer screening: a global overview of existing programmes. *Gut*. 2015; 64: 1637-49.

5. European Council. Council recommendation of 2 December 2003 on cancer screening. http://data.europa.eu/eli/reco/2003/878/oj. Accessed March 23, 2022.

6. McFerran E, O'Mahony JF, Naber S, et al. Colorectal Cancer Screening within Colonoscopy Capacity Constraints: Can FIT-Based Programs Save More Lives by Trading off More Sensitive Test Cutoffs against Longer Screening Intervals? *MDM Policy Pract*. 2022; 7: 23814683221097064.

7. Meklin J, Syrjänen K, Eskelinen M. Colorectal cancer screening with traditional and new-generation fecal immunochemical tests: a critical review of fecal occult blood tests. *Anticancer Res*. 2020; 40: 575-81.

8. Gold MR, Russell LB, Siegel JE, et al. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.

9. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009; 62: e1-e34.

10. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of colorectal cancer screening. *Epidemiol Rev*. 2011; 33: 88-100.

11. Ran T, Cheng C-Y, Misselwitz B, et al. Cost-effectiveness of colorectal cancer screening strategies—a systematic review. *Clin Gastroenterol Hepatol*. 2019; 17: 1969-81. e15.

12. Mendivil J, Appierto M, Aceituno S, et al. Economic evaluations of screening strategies for the early detection of colorectal cancer in the average-risk population: A systematic literature review. *PLoS One*. 2019; 14: e0227251.

13. Zhong G-C, Sun W-P, Wan L, et al. Efficacy and cost-effectiveness of fecal immunochemical test versus colonoscopy in colorectal cancer screening: a systematic review and meta-analysis. *Gastrointest Endosc*. 2020; 91: 684-97. e15.

14. Pignone M, Saha S, Hoerger T, et al. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the US Preventive Services Task Force. *Ann Intern Med*. 2002; 137: 96-104.

15. Patel SS, Kilgore ML. Cost effectiveness of colorectal cancer screening strategies. *Cancer Control*. 2015; 22: 248-58.

16. Jeong KE, Cairns JA. Review of economic evidence in the prevention and early detection of colorectal cancer. *Health Econ Rev*. 2013; 3: 1-10.

17. Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, 2015.

18. Areia M, Fuccio L, Hassan C, et al. Cost-utility analysis of colonoscopy or faecal immunochemical test for population-based organised colorectal cancer screening. *United European Gastroenterol J*. 2019; 7: 105-13.

19. Aronsson M, Carlsson P, Levin L-Å, et al. Cost-effectiveness of high-sensitivity faecal immunochemical test and colonoscopy screening for colorectal cancer. *Br J Surg*. 2017; 104: 1078-86.

20. Arrospide A, Idigoras I, Mar J, et al. Cost-effectiveness and budget impact analyses of a colorectal cancer screening programme in a high adenoma prevalence scenario using MISCAN-Colon microsimulation model. *BMC Cancer*. 2018; 18: 1-11.

21. Babela R, Orsagh A, Ricova J, et al. Cost-effectiveness of colorectal cancer screening in Slovakia. *Eur J Cancer Prev*. 2021; 31: 415-21.

22. Barré S, Leleu H, Benamouzig R, et al. Cost-effectiveness analysis of alternative colon cancer screening strategies in the context of the French national screening program. *Therap Adv Gastroenterol*. 2020; 13: 1756284820953364.

23. Berchi C, Bouvier V, Réaud JM, et al. Cost-effectiveness analysis of two strategies for mass screening for colorectal cancer in France. *Health Econ*. 2004; 13: 227-38.

24. Coretti S, Ruggeri M, Dibidino R, et al. Economic evaluation of colorectal cancer screening programs: Affordability for the health service. *J Med Screen*. 2020; 27: 186-93.

25. Chauvin P, Josselin J-M, Heresbach D. Incremental net benefit and acceptability of alternative health policies: a case study of mass screening for colorectal cancer. *Eur J Health Econ*. 2012; 13: 237-50.

26. Currais P, de Ferro SM, Areia M, et al. Should Colorectal Cancer Screening in Portugal Start at the Age of 45 Years? A Cost-Utility Analysis. *GE Port J Gastroenterol*. 2021; 28: 311-18.

27. Goede SL, van Roon AH, Reijerink JC, et al. Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening. *Gut*. 2013; 62: 727-34.

28. Greuter MJ, Berkhof J, Fijneman RJ, et al. The potential of imaging techniques as a screening tool for colorectal cancer: a cost-effectiveness analysis. *Br J Radiol*. 2016; 89: 20150910.

29. Gyrd-Hansen D. Fecal occult blood tests: a cost-effectiveness analysis. *Int J Technol Assess Health Care*. 1998; 14: 290-301.

30. Gyrd-Hansen D, Søggard J, Kronborg O. Colorectal cancer screening: efficiency and effectiveness. *Health Econ*. 1998; 7: 9-20.

31. Hassan C, Benamouzig R, Spada C, et al. Cost effectiveness and projected national impact of colorectal cancer screening in France. *Endoscopy*. 2011; 43: 780-93.

32. Heresbach D, Chauvin P, Grolier J, et al. Cost-effectiveness of colorectal cancer screening with computed tomography colonography or fecal blood tests. *Eur J Gastroenterol Hepatol*. 2010; 22: 1372-79.

33. Idigoras I, Arrospide A, Portillo I, et al. Evaluation of the colorectal cancer screening Programme in the Basque Country (Spain) and its effectiveness based on the Miscan-colon model. *BMC Public Health*. 2018; 18: 1-12.

34. Jahn B, Sroczynski G, Bundo M, et al. Effectiveness, benefit harm and cost effectiveness of colorectal cancer screening in Austria. *BMC Gastroenterol*. 2019; 19: 1-13.

35. Ladabaum U, Alvarez-Osorio L, Rösch T, et al. Cost-effectiveness of colorectal cancer screening in Germany: current endoscopic and fecal testing strategies versus plasma methylated Septin 9 DNA. *Endosc Int Open*. 2014; 2: E96-E104.

36. Lansdorp-Vogelaar I, Goede SL, Bosch LJ, et al. Cost-effectiveness of high-performance biomarker tests vs fecal immunochemical test for noninvasive colorectal cancer screening. *Clin Gastroenterol Hepatol*. 2018; 16: 504-12. e11.

37. Lejeune C, Arveux P, Dancourt V, et al. Cost-effectiveness analysis of fecal occult blood screening for colorectal cancer. *Int J Technol Assess Health Care*. 2004; 20: 434-39.

38. Lejeune C, Dancourt V, Arveux P, et al. Cost-effectiveness of screening for colorectal cancer in France using a guaiac test versus an immunochemical test. *Int J Technol Assess Health Care*. 2010; 26: 40-47.

39. Lejeune C, Le Gleut K, Cottet V, et al. The cost-effectiveness of immunochemical tests for colorectal cancer screening. *Dig Liver Dis*. 2014; 46: 76-81.

40. Macafee D, Waller M, Whynes D, et al. Population screening for colorectal cancer: the implications of an ageing population. *Br J Cancer*. 2008; 99: 1991-2000.

41. Murphy J, Halloran S, Gray A. Cost-effectiveness of the faecal immunochemical test at a range of positivity thresholds compared with the guaiac faecal occult blood test in the NHS Bowel Cancer Screening Programme in England. *BMJ open*. 2017; 7: e017186.

42. Pil L, Fobelets M, Putman K, et al. Cost-effectiveness and budget impact analysis of a population-based screening program for colorectal cancer. *Eur J Intern Med*. 2016; 32: 72-78.

43. Senore C, Hassan C, Regge D, et al. Cost-effectiveness of colorectal cancer screening programmes using sigmoidoscopy and immunochemical faecal occult blood test. *J Med Screen*. 2019; 26: 76-83.

44. Sharp L, Tilson L, Whyte S, et al. Cost-effectiveness of population-based screening for colorectal cancer: a comparison of guaiac-based faecal occult blood testing, faecal immunochemical testing and flexible sigmoidoscopy. *Br J Cancer*. 2012; 106: 805-16.

45. Sobhani I, Alzahouri K, Ghout I, et al. Cost-effectiveness of mass screening for colorectal cancer: choice of fecal occult blood test and screening strategy. *Dis Colon Rectum*. 2011; 54: 876-86.

46. van Rossum LG, van Rijn AF, Verbeek AL, et al. Colorectal cancer screening comparing no screening, immunochemical and guaiac fecal occult blood tests: A cost-effectiveness analysis. *Int J Cancer*. 2011; 128: 1908-17.

47. Whyte S, Chilcott J, Halloran S. Reappraisal of the options for colorectal cancer screening in England. *Colorectal Dis*. 2012; 14: e547-e61.

48. Wilschut JA, Hol L, Dekker E, et al. Cost-effectiveness analysis of a quantitative immunochemical test for colorectal cancer screening. *Gastroenterology*. 2011; 141: 1648-55. e1.

49. Wilschut JA, Habbema JDF, van Leerdam ME, et al. Fecal occult blood testing when colonoscopy capacity is limited. *J Natl Cancer Inst*. 2011; 103: 1741-51.

50. Thomas C, Mandrik O, Whyte S, et al. Should colorectal cancer screening start at different ages for men and women? Cost-effectiveness analysis for a resource-constrained service. *Cancer Rep*. 2021; 4: e1344.

51. Whyte S, Thomas C, Chilcott J, et al. Optimizing the design of a repeated fecal immunochemical test bowel cancer screening programme with a limited endoscopy capacity from a health economic perspective. *Value Health*. 2022; 25: 954-64.

52. Heisser T, Hoffmeister M, Brenner H. Model based evaluation of long-term efficacy of existing and alternative colorectal cancer screening offers: A case study for Germany. *Int J Cancer*. 2022; 150: 1471-80.

53. van der Meulen MP, Kapidzic A, Leerdam MEv, et al. Do Men and Women Need to Be Screened Differently with Fecal Immunochemical Testing? A Cost-Effectiveness AnalysisFIT Screening Stratified by Gender. *Cancer Epidemiol Biomarkers Prev*. 2017; 26: 1328-36.

54. Tappenden P, Chilcott J, Eggington S, et al. Option appraisal of population-based colorectal cancer screening programmes in England. *Gut*. 2007; 56: 677-84.

55. Heinävaara S, Gini A, Sarkeala T, et al. Optimizing screening with faecal immunochemical test for both sexes-Cost-effectiveness analysis from Finland. *Prev Med*. 2022; 157: 106990.

56. Lee D, Muston D, Sweet A, et al. Cost effectiveness of CT colonography for UK NHS colorectal cancer screening of asymptomatic adults aged 60–69 years. *Appl Health Econ Health Policy*. 2010; 8: 141-54.

57. OECD. Glossary of statistical terms Central and Eastern European Countries (CEECS). https://stats.oecd.org/glossary/detail.asp?ID=303. Accessed November 14, 2022.

58. Meulen MP, Kapidzic A, Leerdam MEv, et al. Do Men and Women Need to Be Screened Differently with Fecal Immunochemical Testing? A Cost-Effectiveness AnalysisFIT Screening Stratified by Gender. *Cancer Epidemiol Biomarkers Prev*. 2017; 26: 1328-36.

59. Heisser T, Weigl K, Hoffmeister M, et al. Age-specific sequence of colorectal cancer screening options in Germany: A model-based critical evaluation. *PLoS Med*. 2020; 17: e1003194.

60. Poskus T, Strupas K, Mikalauskas S, et al. Initial results of the national colorectal cancer screening program in Lithuania. *Eur J Cancer Prev*. 2015; 24: 76-80.

61. Kronborg O, Fenger C, Olsen J, et al. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *The Lancet*. 1996; 348: 1467-71.

62. Hardcastle JD, Chamberlain JO, Robinson MH, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *The Lancet*. 1996; 348: 1472-77.

63. Kewenter J, Brevinge H, Engaras B, et al. Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing: results for 68,308 subjects. *Scand J Gastroenterol*. 1994; 29: 468-73.

64. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med*. 1993; 328: 1365-71.

65. Bénard F, Barkun AN, Martel M, et al. Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations. *World J Gastroenterol*. 2018; 24: 124.

66. Knudsen AB, McMahon PM, Gazelle GS. Use of modeling to evaluate the cost-effectiveness of cancer screening programs. *J Clin Oncol*. 2007; 25: 203-08.

67. Toes-Zoutendijk E, van Leerdam ME, Dekker E, et al. Real-time monitoring of results during first year of Dutch colorectal cancer screening program and optimization by altering fecal immunochemical test cut-off levels. *Gastroenterology*. 2017; 152: 767-75. e2.

68. Lansdorp-Vogelaar I, Meester R, de Jonge L, et al. Risk-stratified strategies in population screening for colorectal cancer. *Int J Cancer*. 2022; 150: 397-405.

Supplementary material 1 The full search strategy.

PubMed

Search concept	Search string	
Colorectal	colon*	[tiab] OR
	rectal	[tiab] OR
	colorectal	[tiab]
Cancer	cancer*	[tiab] OR
	tumour*	[tiab] OR
	tumor*	[tiab] OR
	neoplasm*	[tiab] OR
	carcinoma*	[tiab] OR
	adenoma*	[tiab] OR
	polyp*	[tiab]
Screening	screen*	[tiab] OR
	faecal occult blood	[tiab] OR
	fecal occult blood	[tiab] OR
	*FOBT	[tiab] OR
	faecal immunochemical test	[tiab] OR
	fecal immunochemical test	[tiab] OR
	FIT	[tiab]
Cost-effectiveness filter	cost-effect*	[tiab] OR
	CEA	[tiab] OR
	cost-utility	[tiab] OR
	CUA	[tiab] OR
	health technology assessment	[tiab] OR
	HTA	[tiab] OR
	health economics	[tiab] OR
	cost	[tiab] OR
	cost-benef*	[tiab] OR
	cost anal*	[tiab]
Cost-effectiveness outcomes	quality-adjusted	[tiab] OR
	QALY*	[tiab] OR
	disability-adjusted	[tiab] OR
	DALY*	[tiab] OR
	LYG	[tiab] OR
	life-year*	[tiab] OR
	life expectancy	[tiab]
1 and 2 and 3 and 4 and 5		
Limit to English language and original articles		

Scopus

Search concept	Search string	
Colorectal	TITLE-ABS(	
	colon*	OR
	rectal	OR
	colorectal)	
Cancer	TITLE-ABS(	
	cancer*	OR
	tumour*	OR
	tumor*	OR
	neoplasm*	OR
	carcinoma*	OR
	adenoma*	OR
	polyp*)	
Screening	TITLE-ABS (	
	screen*	OR
	"faecal occult blood"	OR
	"fecal occult blood"	OR
	*FOBT	OR
	"faecal immunochemical test"	OR
	"fecal immunochemical test"	OR
	FIT)	
Cost-effectiveness filter	TITLE-ABS (	
	"cost effect*"	OR
	CEA	OR
	"cost utility"	OR
	CUA	OR
	"health technology assessment"	OR
	HTA	OR
	"health economics"	OR
	costs	OR
	"cost benef*"	OR
	"cost anal*")	
Cost-effectiveness outcomes	TITLE-ABS (	
	"quality adjusted"	OR
	QALY*	OR
	"disability adjusted"	OR
	DALY*	OR
	LYG	OR
	"life year*"	OR
	"life expectancy")	
1 and 2 and 3 and 4		
Limit to English language and original articles		

Web of Science

Search concept	Search string
Colorectal	TS=(
	colon* OR
	rectal OR
	colorectal)
Cancer	TS=(
	cancer* OR
	tumour* OR
	tumor* OR
	neoplasm* OR
	carcinoma* OR
	adenoma* OR
	polyp*)
Screening	TS=(
	screen* OR
	colonoscop* OR
	faecal occult blood OR
	fecal occult blood OR
	*FOBT OR
	faecal immunochemical test OR
	fecal immunochemical test OR
	FIT)
Cost-effectiveness filter	TS=(
	cost effect* OR
	CEA OR
	cost utility OR
	CUA OR
	health technology assessment OR
	HTA OR
	health economics OR
	costs OR
	cost benef* OR
	cost anal*)
Cost-effectiveness outcomes	TS=(
	quality adjusted OR
	QALY* OR
	disability adjusted OR
	DALY* OR
	LYG OR
	life-year* OR
	life expectancy)
1 and 2 and 3 and 4	
Limit to English language and original articles	

Supplementary material 2 Checklist questions (adapted from Drummond et al)

Checklist	
1.	Was a well-defined question posed in answerable form? 1.1 Did the study examine both costs and effects of the service(s) or programme(s) over an appropriate time horizon? 1.2 Did the study involve a comparison of alternatives? 1.3 Was a perspective for the analysis stated and was the study placed in any particular decision-making context? 1.4 Were the patient population and any relevant subgroups adequately defined?
2.	Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)? 2.1 Were any relevant alternatives omitted? 2.2 Was (should) a 'do nothing' alternative (be) considered?
3.	Was the effectiveness of the programme or services established? 3.1 Was this done through a randomized controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice? 3.2 Were effectiveness data collected and summarized through a systematic overview of clinical studies? If so, were the search strategy and rules for inclusion or exclusion outlined? 3.3 Were observational data or assumptions used to establish effectiveness? If so, were any potential biases recognized?
4.	Were all the important and relevant costs and consequences for each alternative identified? 4.2 Did it cover all relevant perspectives? (Possible perspectives include those of patients and third-party payers; other perspectivess may also be relevant depending on the particular analysis.)
5.	Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life years)? 5.1 Were the sources of resource utilization described and justified? 5.2 Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis?
6.	Were the cost and consequences valued credibly? 6.1 Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policymakers' views, and health professionals' judgements.) 6.4 Was the valuation of consequences appropriate for the question posed (i.e. has the appropriate type or types of analysis—cost-effectiveness, cost–benefit—been selected)?
7.	Were costs and consequences adjusted for differential timing? 7.1 Were costs and consequences that occur in the future 'discounted' to their present values? 7.2 Was any justification given for the discount rate(s) used?
8.	Was an incremental analysis of costs and consequences of alternatives performed? 8.1 Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated?
9.	Was allowance made for uncertainty in the estimates of costs and consequences? 9.2 If a sensitivity analysis was employed, was justification provided for the form(s) of sensitivity analysis employed and the ranges or distributions of values (for key study parameters)? 9.3 Were the conclusions of the study sensitive to the uncertainty in the results, as quantified by the statistical and/or sensitivity analysis?
10.	Did the presentation and discussion of study results include all issues of concern to users? 10.1 Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion? 10.2 Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology? 10.4 Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)? 10.5 Did the study discuss issues of implementation, such as the feasibility of adopting the 'preferred' programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes?

10.6 Were the implications of uncertainty for decision-making, including the need for future research, explored?

Sub-questions from the original Drummond checklist that were omitted in our application

2.3 Were relevant alternatives identified for the patient subgroups?

4.1 Was the range wide enough for the research question at hand?

4.3 Were capital costs, as well as operating costs, included?

5.3 Were there any special circumstances (e.g. joint use of resources) that made measurement difficult? Were these circumstances handled appropriately?

6.2 Were market values employed for changes involving resources gained or depleted?

6.3 Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (e.g. clinic space donated at a reduced rate), were adjustments made to approximate market values?

9.1 If patient-level data on costs or consequences were available, were appropriate statistical analyses performed?

9.4 Was heterogeneity in the patient population recognized, for example by presenting study results for relevant subgroups?

10.3 Did the study discuss the generalizability of the results to other settings and patient/client groups?

The detailed results of quality appraisal of studies.

Authors, year of publication	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	TS
Aeria et al, 2019	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	9.5
Aronsson et al, 2017	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	9.0
Arrospide et al, 2018	1.0	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Babela et al, 2021	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	9.0
Barre et al, 2020	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.5	8.5
Berchi et al, 2004	0.5	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Chauvin et al, 2012	1.0	0.5	1.0	0.5	0.5	1.0	0.5	1.0	1.0	0.5	7.5
Coretti et al, 2020	0.5	1.0	1.0	0.0	1.0	1.0	0.5	1.0	1.0	0.5	7.5
Currais et al,2021	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0
Goede et al, 2013	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.5
Greuter et al, 2016	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	8.5
Gyrd Hansen et al, 1998a	1.0	1.0	1.0	0.5	1.0	1.0	0.5	1.0	0.0	0.5	7.5
Gyrd Hansen et al, 1998b	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	9.0
Hassan et al, 2011	1.0	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Heinävaara et al, 2022	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	9.0
Heresbach et al, 2010	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	0.5	0.5	8.5
Jahn et al, 2019	1.0	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Ladabum et al, 2014	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.0
Lansdorp-Vogelaar et al, 2018	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.5
Lee et al, 2010	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0
Leujene et al, 2004	1.0	0.5	1.0	1.0	1.0	0.5	0.5	0.5	1.0	0.5	7.5
Lejeune et al, 2010	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	0.5	0.5	8.5
Lejeune et al, 2014	1.0	1.0	1.0	0.5	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Macafee et al, 2008	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.0	0.5	0.5	7.5
McFerran et al, 2022	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.0
Murphy et al, 2017	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	9.5
Pil et al, 2016	0.5	0.5	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	8.0
Senore et al, 2019	1.0	0.5	0.5	0.5	1.0	1.0	0.5	1.0	1.0	0.5	7.5
Sharp et al, 2012	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.5	9.0
Sobhani et al, 2011	1.0	0.5	1.0	0.5	1.0	1.0	1.0	1.0	0.5	0.5	8.0
Tappenden et al, 2007	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	9.0
Thomas 2021	1.0	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	8.5
Van der Meulen, 2017	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	9.5
Van Rosum et al, 2011	0.5	1.0	1.0	1.0	1.0	0.5	1.0	1.0	1.0	0.5	8.5
Whyte et al, 2012	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.5	0.5	8.0
Whyte et al, 2021	1.0	1.0	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	9.5
Wilschut et al, 2011a	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	9.0
Wilschut 2011b	1.0	1.0	1.0	1.0	1.0	1.0	0.5	1.0	1.0	0.5	9.0

Q, question; TS, total score.

Supplementary material 3 Categorization of aim/objective stated in the studies included in the review

Study	Title	Stated Aim/Objective	Categorisation of the objective
Aeria et al, 2019	Cost-utility analysis of colonoscopy or faecal immunochemical test for population-based organised colorectal cancer screening	"Primary aim of our analysis was to perform a cost utility analysis in a European country, to compare a population-based CRC screening programme with FIT or primary colonoscopy versus no screening. Secondary aims were to explore the influence of adherence rates and costs on such cost-utility ratio, and to evaluate its impact on health resources, namely on the annual capacity for performing colonoscopies."	Broad
Aronsson et al, 2017	Cost-effectiveness of high-sensitivity faecal immunochemical test and colonoscopy screening for colorectal cancer	"The primary aim of the present study was to estimate the long-term costs and health outcomes of four strategies for colorectal cancer screening compared with no screening. A secondary aim was to calculate the uncertainties in the cost-effectiveness results and single-model inputs, such as the importance of the participation rate, age at initiation of screening and disease progression."	Narrow
Arrospide et al, 2018	Cost-effectiveness and budget impact analyses of a colorectal cancer screening programme in a high adenoma prevalence scenario using MISCAN-Colon microsimulation model	"The object of our study was to economically evaluate the programme compared to no screening via cost-effectiveness and budget impact analyses with microsimulation models calibrated to the Basque epidemiological features of CRC incidence and adenoma prevalence."	Broad
Babela et al, 2021	Cost-effectiveness of colorectal cancer screening in Slovakia	"The study aims to analyze the comparative effectiveness and cost-effectiveness of two most often used strategies in Europe and the USA, annual and biennial FIT, in reducing CRC incidence and mortality in the Slovak population of 50–75 years of age using the microsimulation screening analysis model."	Narrow
Barre et al, 2020	Cost-effectiveness analysis of alternative colon cancer screening strategies in the context of the French national screening program	"The objective of this work is to compare the cost effectiveness of screening alternatives currently available for average-risk individuals taking real-world participation rates into account."	Broad
Berchi et al, 2004	Cost-e¡ectiveness analysis of two strategies for mass screening for colorectal cancer in France	"The aim of the present study was to use a simulation model to analyse the costs and the effectiveness of 20 years of biennial CRC screening with an automated immunological test (Magstream) and to compare its results to those obtained with a guaiac stool test (Haemoccult)."	Narrow
Chauvin 2012	Incremental net benefit and acceptability of alternative health policies: a case study of mass screening for colorectal cancer	"Our cost-effectiveness analysis aimed at assessing the potential of alternative mass screening strategies when biennial guaiac FOBT is already implemented in average risk population. We considered three alternative screening policies: biennial immunological FOBT, CTC every 10 years, and CTC every 5 years."	Broad
Coretti et al, 2020	Economic evaluation of colorectal cancer screening programs: Affordability for the health service	"We aimed to investigate the cost-effectiveness of the Abruzzo CRC screening program, and to determine which variables most affect its success. The results may provide a generalizable decisional support tool to steer any revision of the current program."	Narrow
Currais et al, 2021	Should Colorectal Cancer Screening in Portugal Start at the Age of 45 Years? A Cost-Utility Analysis	"We aimed to evaluate if the changes in the CRC incidence/mortality observed in the USA and the rest of Europe also occur in Portugal, and then perform a cost-utility analysis of CRC screening that starts at 45 years of age."	Broad

Goede et al, 2013	Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening	"The aim of this study was to evaluate the cost-effectiveness of one-sample and two-sample FIT screening strategies with variable intervals, age ranges and cut-off levels in order to assess whether the increased performance of a second FIT sample outweighs the increased costs compared with one-sample FIT screening."	Broad
Greuter et al, 2016	The potential of imaging techniques as a screening tool for colorectal cancer: a cost-effectiveness analysis	"We aimed to compare CTC and MRC screening with no screening and colonoscopy screening in terms of effectiveness, costs and cost-effectiveness using the Adenoma and Serrated pathway to Colorectal CAncer (ASCCA) model."	Broad
Gyrd Hansen 1998a	Fecal Occult Blood Tests A cost effectiveness Analysis	"The aim of this paper is to compare the cost-effectiveness of the unhydrated H-II test based on the experience of the Danish randomized trial with the potential cost-effectiveness of alternative FOB tests if these were to be used as a test in a similar population screening program."	Broad
Gyrd Hansen 1998b	Colorectal cancer screening: efficiency and effectiveness	"The aim of this analysis was to build upon the knowledge acquired from this screening programme, in order to give a basis on which to evaluate the cost-effectiveness of alternative colorectal cancer screening programmes using the unhydrated H-II test."	Broad
Hassan et al, 2011	Cost effectiveness and projected national impact of colorectal cancer screening in France	"The aim of the current cost-effectiveness analysis was to assess the efficacy and costs of different CRC screening modalities, including endoscopic modalities, in a simulated cohort of average-risk French subjects, to project the health and economic outcomes on the French population, and to quantify the potential loss associated with decision uncertainty."	Broad
Heinävaara et al, 2022	Optimizing screening with faecal immunochemical test for both sexes - Cost-effectiveness analysis from Finland	"This study aims to model cost-effectiveness of sex-specific strategies for the whole population, and to assess whether the current strategies are optimal."	Broad
Heresbach et al, 2010	Cost-effectiveness of colorectal cancer screening with computed tomography colonography or fecal blood tests	"The aim of our study was thus to assess the cost-effectiveness of CTC and iFOBTstrategies compared with no screening or gFOBT over a 30-year horizon with sex distinction."	Broad
Idigoras et al, 2018	Evaluation of the colorectal cancer screening Programme in the Basque Country (Spain) and its effectiveness based on the Miscan-colon model	"The aim of our study was to determine possible scenarios in terms of incidence, mortality and reduction of Life-years-Lost (L-y-L) in the medium and long term of CRC. The objectives of this study were to predict future scenarios and outcomes for the Basque population and to determine the epidemiological benefits of the screening programme in terms of incidence, mortality and years of life lost (L-y-L)."	Narrow
Jahn et al, 2019	Effectiveness, benefit harm and cost effectiveness of colorectal cancer screening in Austria	"We aim to systematically evaluate the long-term effectiveness, harms and cost effectiveness of different organized CRC screening strategies in Austria."	Broad
Ladabum et al, 2014	Cost-effectiveness of colorectal cancer screening in Germany: current endoscopic and fecal testing strategies versus plasma methylated Septin 9 DNA	"We performed a cost-effectiveness analysis of current and emerging CRC screening strategies in Germany."	Broad
Lansdorp-Vogelaar et al, 2018	Cost-effectiveness of High-performance Biomarker Tests vs Fecal Immunochemical Test for Noninvasive Colorectal Cancer Screening	"Therefore the aim of this study was to provide insight in the requirements for test sensitivity, specificity and unit cost in order for new technologies to be cost-effective compared to FIT screening in population-based screening programs."	Broad

Lee 2010	Cost Effectiveness of CT Colonography for UK NHS Colorectal Cancer Screening of Asymptomatic Adults Aged 60–69 Years	"The current study extends this previous work and uses updated data on disease staging and progression to estimate the cost effectiveness of CRC screening technologies (CTC, OC, FOBT,FSIG) from the perspective of the UK NHS."	Broad
Lejeune et al, 2004	Cost-effectiveness analysis of fecal occult blood screening for colorectal cancer	"The objective of this study was to assess the cost-effectiveness of screening for colorectal cancer using a fecal occult blood screening test, the Hemoccult-II, in a cohort of 100,000 asymptomatic individuals 50–74 years of age."	Narrow
Lejeune et al, 2010	Cost-effectiveness of screening for colorectal cancer in France using a guaiac test versus an immunochemical test	"The aim of this study was to compare the cost and the effectiveness of two biennial fecal occult blood screening tests for colorectal cancer: a guaiac nonrehydrated test (G-FOBT) and an immunochemical test (I-FOBT) with the absence of screening."	Narrow
Lejeune et al, 2014	The cost-effectiveness of immunochemical tests for colorectal cancer screening	"To compare the cost and effectiveness in biennial screening for colorectal cancer of fifteen strategies consisting of the three-stool sample un-rehydrated guaiac faecal occult blood test and three immunochemical tests: Magstream, FOB-Gold and OC-Sensor, at different positivity cut-off levels and stool-sample collection."	Narrow
Macafee et al, 2008	Population screening for colorectal cancer: the implications of an ageing population	"The aim of this study was to simulate a population-based screening setting using a Markov model and assess the effect of increasing life expectancy on CRC screening cost-effectiveness."	Broad
McFerran et al, 2022	Colorectal Cancer Screening within Colonoscopy Capacity Constraints: Can FIT-Based Programs Save More Lives by Trading off More Sensitive Test Cutoffs against Longer Screening Intervals?	"The objective of this study is to further explore the potential for improved effectiveness and cost-effectiveness within a capacity-constrained system."	Broad
Murphy et al, 2017	Cost-effectiveness of the faecal immunochemical test at a range of positivity thresholds compared with the guaiac faecal occult blood test in the NHS Bowel Cancer Screening Programme in England	"The aim of this analysis was to estimate the cost–utility of the faecal immunochemical test for haemoglobin (FIT) compared with gFOBT for a cohort beginning screening aged 60 years at a range of FIT positivity thresholds."	Broad
Pil et al, 2016	Cost-effectiveness and budget impact analysis of a population-based screening program for colorectal cancer	"This study assessed the cost effectiveness and budget impact of the colorectal population-based screening program in Flanders (Belgium)."	Broad
Senore et al, 2019	Cost-effectiveness of colorectal cancer screening programmes using sigmoidoscopy and immunochemical faecal occult blood test	"This analysis aimed to compute the actual cost of organized FS and/or FIT programmes, and to assess their cost-effectiveness, based on real data extracted from an established screening programme."	Broad
Sharp et al, 2012	Cost-effectiveness of population-based screening for colorectal cancer: a comparison of guaiac-based faecal occult blood testing, faecal immunochemical testing and flexible sigmoidoscopy	"The aim of this study was to estimate the incremental cost-effectiveness of a population-based colorectal cancer-screening programme using primary FIT, gFOBT with reflex FIT testing, or once-only FSIG compared with no screening."	Broad
Sobhani et al, 2011	Cost-effectiveness of mass screening for colorectal cancer: choice of fecal occult blood test and screening strategy.	"The aim was to improve the sensitivity of the test by choosing an accurate format (1- to 3-sample and optimum hemoglobin concentration) while maintaining acceptable specificity and avoiding alteration of the screening program in terms of quality of life and economic outputs."	Narrow
Tappenden et al, 2007	Option appraisal of population-based colorectal cancer screening programmes in England	"The aim of this study was to estimate the likely effectiveness, cost-effectiveness and resource impact resulting from the implementation of	Broad

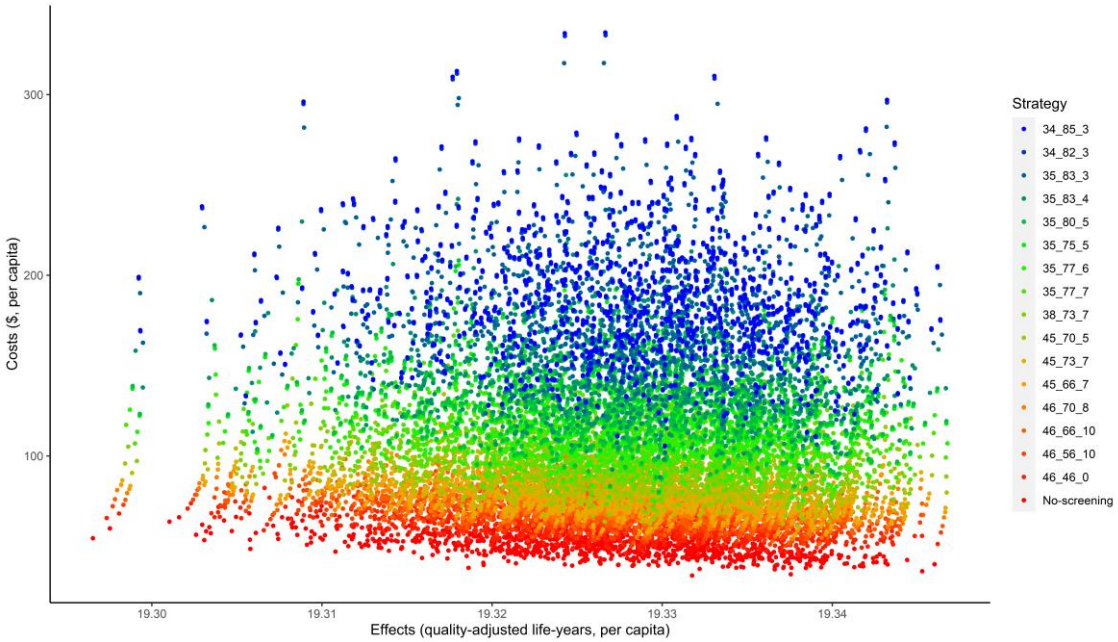
		alternative CRC screening programmes using FOBT, FSIG or a combination of both tests."	
Thomas et al, 2021	Should colorectal cancer screening start at different agesfor men and women? Cost-effectiveness analysis for aresource-constrained service	"This analysis investigates whether, in a resource-constrained setting, it would be more effective and cost-effective for men and women to start screening for CRC at different ages."	Broad
Van der Meulen et al, 2017	Do Men and Women Need to Be Screened Differently with Fecal Immunochemical Testing? A Cost-Effectiveness Analysis	"We used the microsimulation model MISCAN-Colon to evaluate whether screening stratified by gender is cost-effective."	Broad
Van Rossum et al, 2011	Colorectal cancer screening comparing no screening, immunochemical and guaiac fecal occult blood tests: a cost-effectiveness analysis	"We analyzed prospective empirical data from a randomized-controlled trial to compare cost-effectiveness of screening with either one round of immunochemical fecal occult blood testing (I-FOBT; OC-SensorVR), one round of guaiac FOBT (G-FOBT; Hemoccult-IIVR) or no screening in Dutch aged 50 to 75 years, completed with cancer registry and literature data, from a third-party payer perspective in a Markov model with first- and second-order Monte Carlo simulation."	Narrow
Whyte et al, 2012	Reappraisal of the options for colorectal cancer screening in England	"The aim was to use newly available data to estimate the cost effectiveness and endoscopy requirements of screening options for colorectal cancer (CRC) to inform screening policy in England."	Broad
Whyte et al, 2021	Optimizing the Design of a Repeated Fecal Immunochemical Test Bowel Cancer Screening Programme With a Limited Endoscopy Capacity From a Health Economic Perspective	"The aim was to determine the optimal strategy in terms of screening age range, screening frequency, and FIT threshold to inform policy making."	Broad
Wilschut et al, 2011a	Cost-effectiveness Analysis of a Quantitative Immunochemical Test for Colorectal Cancer Screening	"We assessed which are the most clinically effective and cost-effective FOBT screening alternatives under different colonoscopy capacity levels with the validated MISCANColon microsimulation model, using attendance rates, costs, positivity, and detection rates of gFOBT and FIT at varying hemoglobin cutoff levels from two implementation trials in the Netherlands."	Broad
Wilschut et al, 2011b	Fecal Occult Blood Testing When Colonoscopy Capacity is Limited	"We aimed to identify the most cost-effective FIT cut off level for referral to colonoscopy based on data from these trials and allowing for differences in screening ages."	Narrow

APPENDIX 2 PROBABILISTIC SENSITIVITY ANALYSIS

Considering the computational expensive, this thesis limited probabilistic sensitivity analysis (PSA) runs to the strategies situated on the efficient frontier in the base-case (see Table 3-1). Diagram A shows the distribution of these strategies on the cost-effectiveness plane through 1000 iterations. To demonstrate whether the strategies maintain their relative positions on the cost-effectiveness plane, Diagram B connected the strategies across iterations, showing some scenarios have frontiers that are more convex than others. From the results, it is hard to definitively identify which parameter causes this change in convexity. Unlike one-way sensitivity analysis (OWSA), the change in the convexity of the frontiers can be directly linked back to a specific parameter. The discussion of OWSA can be found in the third study of this thesis.

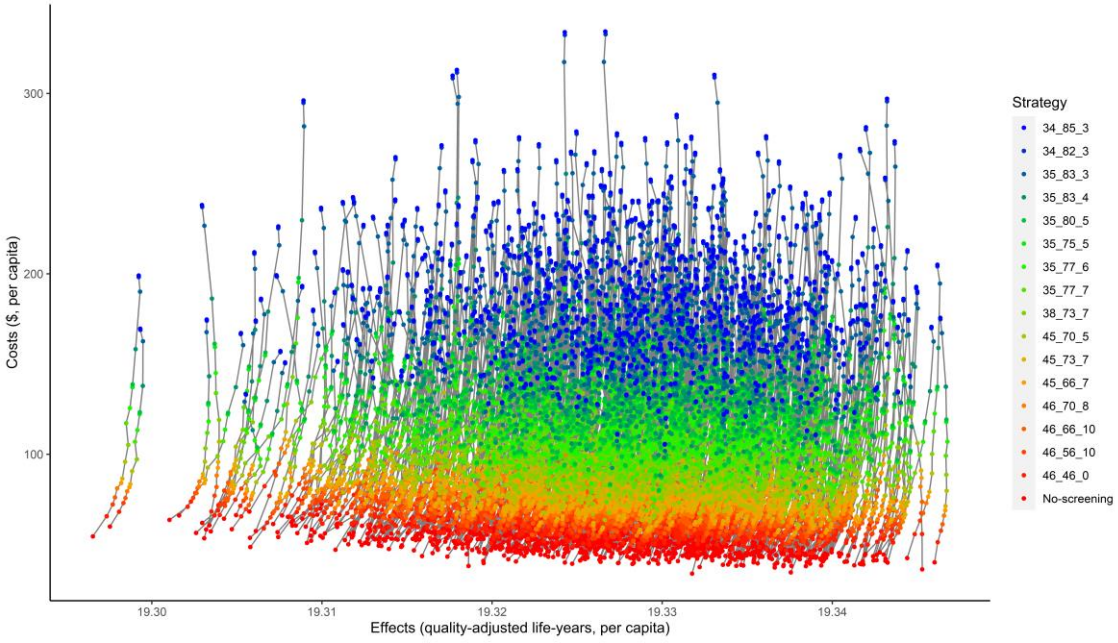
The parameter distribution can be found in Table 1 which are assumed arbitrarily. The parameter values employed do not mirror actual cases, limiting the empirical significance. The chosen parameter set happens to orient the strategy distribution horizontally in Diagram A, but this is purely a hypothetical assumption. Adjusting the parameter distribution, for instance by increasing the uncertainty of cost-relevant parameters, would likely result in a wider array of cost estimates, spreading vertically, according to the observation in OWSA.

Diagram A The cost-effectiveness plane of probabilistic sensitivity analysis results



Only the strategies on the efficient frontier in the base-case scenario are simulated. Each strategy is labelled with the range of screening ages and screening interval.

Diagram B The cost-effectiveness plane of probabilistic sensitivity analysis results (strategies are linked by iteration)



Only the strategies on the efficient frontier in the base-case scenario are simulated. Each strategy is labelled with the range of screening ages and screening interval. Diagram A and B plots the same results, and Diagram B further links the strategies together to demonstrate the relative position of strategies in each iteration.

Table 1 Parameter distribution

Parameter	Distribution (α, β)			Mean	Percentile	
					2.5th	97.5th
Sojourn time						
Preclinical	Weibull	1	5	5.0	0.1265	18.4508
clinical	Uniform*	0.4	0.6	0.5	0.4100	0.5900
Health state utility**						
Disease free	Beta	9990	10	0.999	0.9984	0.9995
Preclinical	Beta	5	45	0.1	0.0411	0.1771
Clinical	Beta	40	60	0.4	0.3209	0.4814
Screening modality						
Test sensitivity	Beta	38	2	0.95	0.8840	0.9908
Test specificity	Beta	49	1	0.98	0.9407	0.9990
Treatment effectiveness						
Preclinical	Beta	425	75	0.85	0.8230	0.8754
Clinical	Beta	200	300	0.40	0.3642	0.4362
Intervention disutility						
Primary screen	Beta	1	999	0.001	0.0001	0.0030
Triage	Beta	6	594	0.01	0.0044	0.0175
Treatment	Beta	45	455	0.09	0.0699	0.1119
Screen cost						
Primary screen	Gamma	25	4	100	70	135
Follow-up	Gamma	250	4	1,000	898	1,106
Treatment cost						
Screen-detected	Gamma	1,000	10	10,000	9,486	10,526
Symptomatic	Gamma	1,000	20	20,000	18,971	21,052
Incidence						
Age 30-45	Beta	5	9,995	0.0005	0.0002	0.0009
Age 45-60	Beta	16	19,984	0.0008	0.0005	0.0012
Age 60-80	Beta	24	19,976	0.0012	0.0008	0.0016
Over age 80	Beta	16	9,984	0.0016	0.0010	0.0023

All the values are defined arbitrarily, solely for the purpose of demonstrating the pedagogical model.

*For the Uniform distribution, α and β separately represent the minimum and maximum values.

**To emphasize the importance of early health states in the model, we modelled the decrement in utility for preclinical and clinical health states. For example, the utility of clinical state 0.499 is calculated as 0.999 (disease-free) - 0.1 (preclinical) - 0.4 (clinical).

APPENDIX 3 SIMULATION CODE

1. Firstly remove the saved data in the memory of R software and save the current time for measuring the simulation time.

```
rm(list = ls())  # Delete everything that is in R's memory
time <- Sys.time()  # Save system time
```

2. Set up the working directory to let the software know where is the folder we are using. The last line of code is for automatically setting the working directory to the file location.

```
# Set the working file directory path. Paste your working directory
within the quotes
# Ensure that the backslashes \ are changed to forward slashes /
# You can put as many file paths in as you like here, R will take the
last one that worked
# The try function simply masks error reporting: it will work if one
of the directories attempted is correct

try(setwd("Your working directory 1"), silent = TRUE)
try(setwd("Your working directory 2"), silent = TRUE)
try(setwd(dirname(sys.frame(1)$ofile)), silent = TRUE)  # Drive file
path
try(setwd(dirname(rstudioapi::getActiveDocumentContext()$path)),
silent = TRUE)  # Drive file path
```

3. Import saved parameter inputs by reading the files in *InputFiles* folder.

```
# Import the required data from the data sheets created by the Excel
file
InputTable       <- data.frame(read.table(file =
"InputFiles/InputTable.txt",       header = T, row.names = 1))  #
Apply the first column as row names
DefineStages     <- data.frame(read.table(file =
"InputFiles/Stages.txt",           header = T))  # Stage;
Name;Type;Scale;Shape;Utility
Incidence        <- data.frame(read.table(file =
"InputFiles/Incidence.txt",        header = T))  # Age_Interval;
Annual_Incidence; Cumulative_Failure
Survival         <- data.frame(read.table(file =
"InputFiles/LifeTable.txt",        header = T))  # Interval_LifeTable;
Survival
ScreenSchedule   <- data.frame(read.table(file =
"InputFiles/ScreenSchedule.txt",   header = T))  # ScheduleNumber;
```

```
StartAge; StopAge; TestApplied1; Interval1
TestPerformance  <- data.frame(read.table(file =
"InputFiles/TestPerformance.txt",  header = T))  # Test; Sn; Sp;
Disutility
TreatmentSuccess <- data.frame(read.table(file =
"InputFiles/TreatmentSuccess.txt", header = T))  #
PreClinicalProbability; ClinicalProbability
Disutility       <- data.frame(read.table(file =
"InputFiles/OtherDisutility.txt",  header = T))  #
PreClinicalProbability; ClinicalProbability
DiscountRates    <- data.frame(read.table(file =
"InputFiles/DiscountRates.txt",    header = T))  # Costs; Effects;
DiscountYear
Costs            <- data.frame(read.table(file =
"InputFiles/Costs.txt",            header = T))  # PrimaryScreen;
FollowUp; TreatmentScreenDetected; TreatmentClinical
source("InputFiles/Misc.txt")  # SampleSize; Threshold
source("InputFiles/Options.txt")
```

4. Prepare an Input table that has the parameter values for every scenario. First, make a list of parameters that we want to simulate for one-way sensitivity analysis. Remove the symbol # in front of the parameter to include this parameter for sensitivity analysis.

Then, create a data frame Input from InputTable which the first column records the parameter values for the base-case scenario. InputTable is a table recorded all the parameter input values in Excel.

Create a string to save the names of scenarios which will be used as column names of Input table. We save Input, the name of the first column of this table, to the variable ScenarioNames.

```
# To make a list of parameters that are included in scenario analysis
ParameterNames <- c("StageScale2",
                    "StageScale3",
                    #"StageUtility1",  # These parameters are not
included in our scenario analysis
                    #"StageUtility2",
                    #"StageUtility3",
                    "TestSensitivity1",
                    "TestSpecificity1",
                    #"TestDisutility1",
                    "PreClinicalProbability",
                    "ClinicalProbability",
                    #"DisutilityTriage",
                    #"DisutilityTrt",
```

```
                    #"DiscountRateCost",
                    #"DiscountRateEffect",
                    #"DiscountYear",
                    "CostPrimaryScreen",
                    "CostFollowUp",
                    "CostTrtScreen",
                    "CostTrtClinical",
                    "Incidence",  # In this example, we change the
value of incidence and survival proportionally across age groups
                    "Survival"
                    )

Input <- data.frame(InputTable[, "Input"], row.names =
rownames(InputTable))  # Create a table to record parameter inputs
ScenarioNames <- "Input"  # Create a variable to save the names of
scenarios
```

5. Prepare the parameters' input for each scenario if the user intends to simulate the one-way sensitivity analysis. Each parameter has two values separately for the scenario of low value and the scenario of high value. Each scenario only changes one parameter at a time and the rest parameters remain the fixed default values as the base-case scenario. Two loops are created for both parameters and scenarios. ParameterNames are defined in step 4 and ScenarioValue includes "Low" and "High". Because incidence and survival are defined by a series of parameters, symbolling different age groups, we have the parameters associated with these two parameters change simultaneously for one scenario. In the end of the loop, assign the variable ScenarioNames to the table Input.

```
if (ScenarioAnaylsis == TRUE){
  ScenarioValue <- c("Low", "High")  # Our default has two scenarios
for each parameter
  for (Parameter in ParameterNames){  # The list of parameters for
scenario analysis
    for (Scenarios in ScenarioValue){
      InputValue <- InputTable[, "Input"]
      if (!(Parameter %in% c("Incidence", "Survival"))){  #
Parameters, incidence and survival, have different risks for age
groups
        InputValue[rownames(InputTable) %in% Parameter] <-
InputTable[rownames(InputTable) %in% Parameter, Scenarios]
      } else if (Parameter == "Incidence") {
          InputValue[rownames(InputTable) %in% paste("Incidence",
Incidence$AgeInterval, sep = "_")] <-
          InputTable[rownames(InputTable) %in% paste("Incidence",
Incidence$AgeInterval, sep = "_"), Scenarios]
        } else {
            InputValue[rownames(InputTable) %in% paste("Survival",
Survival$AgeInterval, sep = "_")] <-
            InputTable[rownames(InputTable) %in% paste("Survival",
Survival$AgeInterval, sep = "_"), Scenarios]
```

```
          }
      Input <- cbind(Input, InputValue)  # Save parameter value to the
input table
      ScenarioNames <- c(ScenarioNames, paste(Parameter, Scenarios,
sep = "_"))
    }
  }
}

colnames(Input) <- ScenarioNames
```

6. Define the screen schedules. If the user intends to simulate the screening schedules defined in Excel, we saved the relevant parameters into strings for later simulation. If the user wants to systematically define the schedules, run the code below to produce all the possible screening strategies. The default is screening starting at age 25 and stop at age 100 with a fixed interval of 1-10 years, which includes the one-time screening within this age range. The first modality defined in Excel remains the same, as the only screening modality applied in the screening programmes. Create a list to record all the strategy names.

```
# We are able to redefine the screen schedule
if (ExcelDefinedScreening == TRUE){
  StartAges           <- ScreenSchedule[, "StartAge"]
  StopAges            <- ScreenSchedule[, "StopAge"]
  IntervalSwitchAge1s <- ScreenSchedule[, "IntervalSwitchAge1"]
  IntervalSwitchAge2s <- ScreenSchedule[, "IntervalSwitchAge2"]
  IntervalSwitchAge3s <- ScreenSchedule[, "IntervalSwitchAge3"]
  Interval1s          <- ScreenSchedule[, "Interval1"]
  Interval2s          <- ScreenSchedule[, "Interval2"]
  Interval3s          <- ScreenSchedule[, "Interval3"]
  Interval4s          <- ScreenSchedule[, "Interval4"]
  TestSwitchAges      <- ScreenSchedule[, "ScreenSwitchAge"]
  ScreenTest1s        <- ScreenSchedule[, "TestApplied1"]
  ScreenTest2s        <- ScreenSchedule[, "TestApplied2"]
} else {
    # Simulate all the possible screening strategies, assuming the
screen interval remains fixed throughout the programme
    Start <- 25  # The starting age of screening
    Stop <- 100  # The stop age of screening
    Frequency <- c(1:10)  # The screen interval is an integer in the
range of 1-10 years

    # One-time screen in the lifetime
    StartAges <- StopAges <- c(seq(Start, Stop, by = 1))
    Interval1s <- rep(0, (Stop - Start + 1))  # No screening interval
    ScreenTest1s <- rep(1, (Stop - Start + 1))  # We assume all the
strategies use the first screening modality
    IntervalSwitchAge1s <- IntervalSwitchAge2s <- IntervalSwitchAge3s
<-
    Interval2s          <- Interval3s          <- Interval4s
```

```
<-
    TestSwitchAges      <-  ScreenTest2s       <- rep(NA, (100 - 25 +
1))    # No screening interval and other screening modality

    # The example here only changes starting age, stop age, and
screening intervals
    for (i in c(seq(Start ,Stop, by = 1))){
      for (j in c(seq(Start, Stop, by = 1))){
        for (k in Frequency){
          if ((i < j) & ((j - i) %% k == 0)){  # Remove non-integer
screens
            StartAges           <- c(StartAges, i)
            StopAges            <- c(StopAges, j)
            IntervalSwitchAge1s <- c(IntervalSwitchAge1s, NA)
            IntervalSwitchAge2s <- c(IntervalSwitchAge2s, NA)
            IntervalSwitchAge3s <- c(IntervalSwitchAge3s, NA)
            Interval1s          <- c(Interval1s, k)
            Interval2s          <- c(Interval2s, NA)
            Interval3s          <- c(Interval3s, NA)
            Interval4s          <- c(Interval4s, NA)
            TestSwitchAges      <- c(TestSwitchAges, NA)
            ScreenTest1s        <- c(ScreenTest1s, 1)  # We assume all
the strategies use the same screening modality
            ScreenTest2s        <- c(ScreenTest2s, NA)
          }
        }
      }
    }
}

# We give strategies shorter names
strategies <- paste(StartAges, StopAges, Interval1s, sep = "_")
#strategies <- paste(StartAges, StopAges, Interval1s,
IntervalSwitchAge1s, IntervalSwitchAge2s, IntervalSwitchAge3s,
#                    Interval2s, Interval3s, Interval4s,
TestSwitchAges, ScreenTest1s, ScreenTest2s, sep = "_")
```

7. Run all the scenarios with a loop of all the input values defined in `Input`. Each column has a set of parameter values for one simulation.

```
for (CurrentRun in c(1: ncol(Input))){
```

8. Have a fixed random seed for the simulation, so we can have identical output when we want to compare results. The number is changeable but needs to be the same if intending to produce constant outputs.

```
  set.seed(2021)  # Set a seed to be able to reproduce the same
results
```

9. Create an outcome table to record the disease history for all the individuals. Our model in default has five health states. With an additional column to record individuals' ID numbers, create an array `Outcomes` of six columns and the number of rows equals the length of the sample size.

```
  # Define number of health stages with the stage arrival
  # This model only features five states: (1)Disease Free;
(2)Preclinical Disease; (3)Clinical Disease; (4)Cause-Specific Death;
(5)Other-Cause Death
  Outcomes <- array(NA, dim = c(SampleSize, 6))  # Create an array of
the length of the sample size
  colnames(Outcomes) <- c("PersonNumber",
paste(DefineStages[2:nrow(DefineStages), "Name"]), "AllCauseDeath")  #
Set column names
  Outcomes[, "PersonNumber"] <- c(1:SampleSize)  # Set the first
column to be the unique person-number for each individual
```

10. Create a function, `OnsetFunction`, to produce the sampled age of entering a specific health state for each individual from an age-specific probability. This is based on linear interpolation that is useful for filling the gaps with the limited known data points on the curve. It will be applied to simulate the age of entering the preclinical state and other-cause death based on the probabilities for different age groups.

```
  # The intervening columns correspond to the arrival of the
intermediate disease states
  # This model studies the cohort with the same age, so here does not
need to simulate the "Stage1Arrival"
  # Define the generic onset function which applies an age-specific
probability of entering a specific stage
  OnsetFunction <- function(x){
    unlist(approx(probability, age, x, ties = max)[2], use.names = F)
# Ties = max is required because of the possibility of multiple zero
probabilities of disease at younger ages
  }
```

11. Calculate the cumulative probability of being disease-free. To achieve this, we set the initial `CumulativeFailure` to one, meaning nobody gets the disease. Then, run a loop to

calculate the cumulative probabilities by the previous probability of having the disease-free minus the probability of getting disease within that age range. The probability of getting the disease within that age range is the annual probability of getting the disease times by the length of years for a specific age group.

```
  CumulativeFailure <- 1
  IncidenceTable <- data.frame(Incidence$AgeInterval,
                               as.numeric(Input[paste("Incidence",
Incidence$AgeInterval, sep = "_"), CurrentRun]),
                               CumulativeFailure
                               )
  colnames(IncidenceTable) <- c("AgeInterval", "AnnualIncidence",
"CumulativeFailure")
  for (c in c(2:nrow(IncidenceTable))){
    IncidenceTable[c, "CumulativeFailure"] <- (IncidenceTable[c - 1,
"CumulativeFailure"] - (IncidenceTable[c, "AnnualIncidence"] *
(IncidenceTable[c, "AgeInterval"] - IncidenceTable[c - 1,
"AgeInterval"])))
  }
```

12. To run the **OnsetFunction**, remove the constant probabilities for the age groups in their early age or late age. In our example, there is no incidence before age 30, so simulating the probability for age 30 is meaningless and is removed from the simulation.

```
  # Define the specific values of probability and age for disease
onset
  probability <- IncidenceTable$CumulativeFailure
  age         <- IncidenceTable$AgeInterval
  # The probability distribution need to be curtailed at the top and
bottom for non-unique probability for the approx function to work as
intended
  age         <- age[max(which(probability ==
max(probability))):length(probability)]  # Remove the lower bound
values
  probability <- probability[max(which(probability ==
max(probability))):length(probability)]  # Remove the lower bound
values
  age         <- age[1:min(which(probability == min(probability)))]  #
Remove the upper bound values
  probability <- probability[1:min(which(probability ==
min(probability)))]  # Remove the upper bound values
```

13. Simulate the age of entering the preclinical health state based on the incidence of different age groups. The variable `X` saves the random values between 0 and 1 for each

individual. The age of entering preclinical state is calculated based on the random value and its corresponding age with the self-defined function **OnsetFunction**.

```
  x           <- runif(SampleSize)  # Create a vector of random values
  Outcomes[, "Preclinical"] <- OnsetFunction(x)  # Find the age of
disease onset by applying the general function
```

14. Simulate the age of entering the other-cause death based on survival (life-table) by accommodating step 12 and step 13. Survival is already defined as the cumulative probabilities, so there is no need to recalculate it, but we need to remove the constant probabilities in the early age and the late age (same as step 12). Run step 13 but save the results to the column named OtherCauseDeath in the table Outcomes.

```
  # Define the specific values of probability and age for other cause
death
  # The intervals in the age-specific incidence are defined by those
used in the life table
  probability <- as.numeric(Input[paste("Survival",
Survival$AgeInterval, sep = "_"), CurrentRun])
  age         <- Survival$AgeInterval
  # The probability distribution need to be curtailed at the top and
bottom for non-unique probability for the approx function to work as
intended
  age         <- age[max(which(probability ==
max(probability))):length(probability)]  # Remove the lower bound
values
  probability <- probability[max(which(probability ==
max(probability))):length(probability)]  # Remove the lower bound
values
  age         <- age[1:min(which(probability == min(probability)))]  #
Remove the upper bound values
  probability <- probability[1:min(which(probability ==
min(probability)))]  # Remove the upper bound values
  x           <- runif(SampleSize)  # Create a vector of random values
  Outcomes[, "OtherCauseDeath"] <- OnsetFunction(x)  # Find the age at
other-cause death by applying the generic onset function
```

15. Simulate the ages of entering other health states except for preclinical and other-cause death. The age of entering the health state is the age of entering the previous state plus the sojourn time of this previous state. Apply a loop for the number of stages minus one to exclude the final stage of other-cause death. Inside the loop, we create random numbers again for drawing the corresponding duration from the defined distributions.

Distributions, including constant, exponential, and Weibull, can be chosen to sample the sojourn time of each health state. Notably, constant duration does not need a random sample because it assumes every individual has a fixed length of duration. After simulating the ages of entering health states, check if the ages are over 100 which is out of our simulation scope. This study assumes that all the people die before 100 years old and 100 is the maximum age of entering any health state in this simulation.

```
  # The model needs a loop here to go through the disease stages
  for (Stage in 1:(nrow(DefineStages) - 1)){
    # Apply the sojourn time to the stages
    # Retrieve the distribution type, scale and shape
    DurationType <- Input[paste("StageType", Stage, sep = ""),
CurrentRun]
    DurationScale <- Input[paste("StageScale", Stage, sep = ""),
CurrentRun]
    DurationShape <- Input[paste("StageShape", Stage, sep = ""),
CurrentRun]

    if (!(is.na(DurationType))){
      if (DurationType == 1){Duration <- rep(DurationScale,
SampleSize)}  # Set the preclinical distribution to be constant
      if (DurationType == 2){Duration <- -(log(1 - runif(SampleSize)))
* DurationScale}  # Set the preclinical distribution to be
exponentially distributed
      if (DurationType == 3){Duration <- ((-log(1 -
runif(SampleSize))) ^ (1 / DurationShape)) * DurationScale}  # Set the
preclinical distribution to be Weibull distribution, and the
alternative code: rweibull(SampleSize, shape = DurationShape, scale =
DurationScale)
      Outcomes[, Stage + 1] <- Outcomes[, Stage] + Duration  # Find
the end of the preclinical period by adding the onset to the duration
    }

    # This study assumes all the people died before aged 100
    Outcomes[which(Outcomes[, Stage + 1] > 100), Stage + 1] <- 100
  }
```

16. Define AllCauseDeath by comparing an individual's cause-specific death and other-cause death. The lower age is the age of all-cause death. Both two health states, cause-specific death and other-cause death, are absorbing states, meaning individuals cannot enter other states after they enter either of them.

```
  # Find the all-cause death
  Outcomes[, "AllCauseDeath"] <- pmin(Outcomes[,
"CauseSpecificDeath"], Outcomes[, "OtherCauseDeath"], na.rm = TRUE)
```

17. Find individuals who are disease-free, sick, and clinically diagnosed in their disease history. Healthy individuals are those who never develop disease or those who develop disease after their other-cause deaths. Sick individuals are the complete cohort excluding the healthy individuals. Clinical cases are the sick individuals clinically presented before they die from other-cause.

```
  # Identify individuals presenting with clinical disease
  DiseaseFree <- which((Outcomes[, "Preclinical"] >= Outcomes[,
"OtherCauseDeath"]) | (is.na(Outcomes[, "Preclinical"])))  # Censor
onset ages greater than death; first identify those who never develop
disease
  Sick <- Outcomes[-DiseaseFree, "PersonNumber"]
  ClinicalCases <- Sick[which(Outcomes[Sick, "Clinical"] <=
Outcomes[Sick, "OtherCauseDeath"])]
```

18. Find cured cases and redefine their age of all-cause death. Create sample numbers for each individual, `CureSeed`, and then compare `CureSeed` to the treatment success probability. `CuredCases` are clinical cases whose sample number is lower than the treatment success probability. For those cured cases, they die from other-cause, instead of cause-specific death. Adjust the all-cause death in our `Outcomes` table for the cured cases.

```
  # Identify individuals cured successfully
  CureSeed <- runif(SampleSize)  # Save random numbers for treatment
  CuredCases <- ClinicalCases[which(CureSeed[ClinicalCases] <=
Input["ClinicalProbability", CurrentRun])]  # Find treatment to be
successful if probabilities less than probability of successful
treatment
  Outcomes[CuredCases, "AllCauseDeath"] <- Outcomes[CuredCases,
"OtherCauseDeath"]
```

19. Create two series of discounting factors for costs and effectiveness respectively. The discount year is also considered. For effectiveness, we create a variable `Acc_DF_Effects` for recording the cumulative discounted life-years. We discount the results on a discrete basis. The beginning of the simulation has the discounting factor equal to one (without discounting). A fixed discounting factor is applied to everything that happens throughout the year and the discounting factor changes at the end of every year.

```
  # Create a reference series of discount factors for 100 years
  DF_Effects <- round((1 + Input["DiscountRateEffect", CurrentRun]) ^
-(c(1:100) - Input["DiscountYear", CurrentRun]), digits = 4)
  DF_Costs   <- round((1 + Input["DiscountRateCost",   CurrentRun]) ^
-(c(1:100) - Input["DiscountYear", CurrentRun]), digits = 4)
  Acc_DF_Effects <- NA  # Create a variable to save accumulated
discounted life-years
  for (year in c(1:100)){
    Acc_DF_Effects <- c(Acc_DF_Effects, sum(DF_Effects[1:year]))
  }
  Acc_DF_Effects <- Acc_DF_Effects[-1]  # Remove NA
```

20. Create a self-defined function PresentValue to calculate the present value of effectiveness (life-years). The discounted life-years are calculated by referencing the integer part of the life-years to the accumulated discounted factors and the rest decimal numbers times to the discounting factor of the next year.

The life-years shorter than one year will be discounted by the first discount factor. If the input number is NA, then output NA.

```
  # Define the present value function for effectiveness
  PresentValue <- function(x){
    if (x >= 1 & !(is.na(x))){
      Acc_DF_Effects[trunc(x)] + (x - trunc(x)) * DF_Effects[trunc(x)
+ 1]
    } else if (!(is.na(x))){
      x * DF_Effects[1]
      } else {
        x
        }
  }
```

21. Prepare for calculating quality-of-life adjusted outcomes. First, we prepare the discount and undiscounted life-years. The undiscounted life-years are the individuals' age of all-cause deaths. The discounted life-years are the applying the PresentValue function to the undiscounted life-years. We create SickOutcome which only sick people included, because sick people experience more complicated quality-of-life adjustment by entering other health states. Similarly, Dis_SickOutcome is created for discounting results.

```
  # Calculate the effects
  UD_QALYS <- Outcomes[, "AllCauseDeath"]
  D_QALYS <- D_LYS <- sapply(UD_QALYS, PresentValue)
  StagesInOrder <- c(paste(DefineStages[2:(nrow(DefineStages) - 2),
"Name"]), "AllCauseDeath")
  SickOutcome <- as.data.frame(Outcomes[Sick, StagesInOrder])  #
Create a table that only sick people included
  Dis_SickOutcome <- apply(SickOutcome, MARGIN = c(1, 2),
PresentValue)  # Discount life-years for each health state among sick
people
  Utility <- Input[paste("StageUtility", c(1:(nrow(DefineStages) -
2)), sep = ""), CurrentRun]  # Quality-of-life utility weights for
each state
```

22. Create a self-defined function `Accumulated_QALYS` to calculate the sum of quality-adjusted life-years (QALYs) of all the health states. QALYs are calculated by the sojourn time of a specific health state times to its corresponding utility weight. Then sum up all the QALYs from every health state. The QALYs from the first health state are the duration of being disease-free times by the utility weight of the disease-free. The duration of being disease-free is the age of entering the preclinical state. Then, create a loop for the rest health states.

```
  # Define the function for quality-of-life adjustment
  Accumulated_QALYS <- function(x, output){
    QALYS <- x[1] * Utility[1]  # For the first health state, simply
multiply the life-years by the corresponding utility weight
    for (state in (2:length(StagesInOrder))){  # For the remaining
health states
        QALYS <- QALYS + Utility[state] * (x[state] - x[state - 1])  #
Add up the QALYs in each health state
        if (state == length(StagesInOrder)){  # The last health state
          break
        }
    }
    return(QALYS)
  }
```

23. Apply the `Accumulated_QALYS` function to the sick people, including the undiscounted results, `UD_QALYS`, and discounted results, `D_QALYS`. For those who are not sick, adjust their QALYs with the utility weight of the healthy state.

```
  UD_QALYS[DiseaseFree] <- UD_QALYS[DiseaseFree] * Utility[1]  #
Quality-of-life adjusted but not discounted
  UD_QALYS[Sick] <- apply(SickOutcome, 1, Accumulated_QALYS)
  D_QALYS[DiseaseFree] <- D_QALYS[DiseaseFree] * Utility[1]  #
Quality-of-life adjusted and discounted
  D_QALYS[Sick] <- apply(Dis_SickOutcome, 1, Accumulated_QALYS)
```

24. Calculate the dis-utility due to the treatment. Only clinical cases receive treatments and we assumed they received the treatment at the end of the year. In the case of calculating undiscounted QALYs loss due to treatment, simply multiply the number of clinical cases by the dis-utility of each treatment. For the calculation of discounted QALYs loss, every dis-utility is discounted by the corresponding discounting factor.

```
  # Dis-utility due to the treatment
  ClinicalOnsetAges <- ceiling(Outcomes[ClinicalCases, "Clinical"])  #
Identify the clinical ages and round up to integers for discounting
  UD_TreatmentHarm <- length(ClinicalCases) * Input["DisutilityTrt",
CurrentRun]  # Undiscounted QALYs loss due to treatment
  D_TreatmentHarm <- sum(Input["DisutilityTrt", CurrentRun] *
DF_Effects[ClinicalOnsetAges])  # Discounted QALYs loss due to
treatment
```

25. Sum up the life-years and QALYs of each individual, including discounted and undiscounted. The calculation of QALYs for the no-screening scenario needs to deduct the dis-utility due to treatment.

```
  # Calculate the effectiveness outcomes without screening
  UD_NoScreening_LYG <- sum(Outcomes[, "AllCauseDeath"])
  D_NoScreening_LYG <- sum(D_LYS)
  UD_NoScreening_QALY <- sum(UD_QALYS) - UD_TreatmentHarm
  D_NoScreening_QALY <- sum(D_QALYS) - D_TreatmentHarm
```

26. Estimate the cost of the no-screening scenario. In this study, we only consider the treatment cost, `LateTreatmentCost`, for the no-screening strategy which cost is applied to the patients presented clinically. Undiscounted cost is the number of clinical cases multiplied by the unit cost of each treatment. Discounted cost is the sum of the product of the unit cost of each treatment and the corresponding discounting factor.

```
  # Retrieve the costs parameters
  ScreenCost <- Input["CostPrimaryScreen", CurrentRun]
  TriageCost <- Input["CostFollowUp", CurrentRun]
  EarlyTreatmentCost <- Input["CostTrtScreen", CurrentRun]
  LateTreatmentCost <- Input["CostTrtClinical", CurrentRun]

  # Calculate the costs of treating clinical disease under no
screening
  UD_NoScreening_Costs <- length(ClinicalCases) * LateTreatmentCost
  D_NoScreening_Costs <- sum(DF_Costs[ClinicalOnsetAges] *
LateTreatmentCost)  # Find the discounted treatment costs
```

27. Calculate intermediate outcomes, including the average preclinical sojourn time, the number of cancer deaths, over-diagnosis, and clinical cases. The preclinical duration is the difference in the age of entering the clinical state and preclinical state. The number of cancer deaths is calculated by identifying whose all-cause death is equal to cause-specific death. Due to no-screening, there is no case of being over-diagnosed. The number of clinical cases is the length of the string `ClinicalCases` which records all the IDs of clinical cases.

```
  # Calculate intermediate outcomes: effects
  PreClinicalDuration <- mean(Outcomes[ClinicalCases, "Clinical"] -
Outcomes[ClinicalCases, "Preclinical"])  # Average preclinical sojourn
time
  N_CancerDeaths <- length(which(Outcomes[, "AllCauseDeath"] ==
Outcomes[, "CauseSpecificDeath"]))  # The number of cases that die at
the age of disease death
  N_Overdiagnosed <- 0
  N_ClinicalCases <- length(ClinicalCases)  # The number of cases that
present clinically
```

28. Create a data frame `CostEffectivenessResults` to record the main cost-effectiveness outcomes.

```
  # Record cost-effectiveness results
  CostEffectivenessResults <- data.frame(StrategyName =
c("NoScreening", strategies),
                                         UD_EffectsL   =
as.numeric(UD_NoScreening_LYG),
                                         UD_EffectsQ   =
as.numeric(UD_NoScreening_QALY),
                                         UD_Costs      =
as.numeric(UD_NoScreening_Costs),
                                         D_EffectsL    =
```

```
as.numeric(D_NoScreening_LYG),
                                        D_EffectsQ    =
as.numeric(D_NoScreening_QALY),
                                        D_Costs       =
as.numeric(D_NoScreening_Costs),
                                        UD_ICER_L     =
as.numeric(NA),
                                        UD_ICER_Q     =
as.numeric(NA),
                                        D_ICER_L      =
as.numeric(NA),
                                        D_ICER_Q      = as.numeric(NA)
                                        )
```

29. Create a self-defined function, AgeDistribution, to track how individuals move among health states every year. Apply this function to the Outcomes table which records individuals' disease history.

```
  # Record age distribution of each health state
  AgeDistribution<- function(x, output){
    Age.cut <- cut(x, seq(0, 100, by = 1), right = FALSE)
    return(table(Age.cut))
  }
  IntermediateOutcomes_AgeDist <- apply(Outcomes[,
!(colnames(Outcomes) %in% "PersonNumber")], 2, AgeDistribution)
```

30. Save the table IntermediateOutcomes_AgeDist to the following folder.

```
  write.table(IntermediateOutcomes_AgeDist,

paste("OutputFiles/LatestResults/IntermediateOutcomtes/AgeDistribution
/", colnames(Input)[CurrentRun], "_NoScreening", ".txt", sep = ""),
              row.names = F, col.names = T, sep = '\t')
```

31. Create a data frame IntermediateOutcomes to record the intermediate outcomes of each strategy. We have the results of the no-screening saved in the first row. The rest rows will be updated with the simulation outputs below.

```
  # Record intermediate outcomes
  IntermediateOutcomes <- data.frame(StrategyName           =
c("NoScreening", strategies),
```

```
                                    UD_ScreenCosts          =
as.numeric(0),
                                    UD_TriageCosts          =
as.numeric(0),
                                    UD_EarlyTreatmentCosts  =
as.numeric(0),
                                    UD_LateTreatmentCosts   =
as.numeric(UD_NoScreening_Costs),
                                    D_ScreenCosts           =
as.numeric(0),
                                    D_TriageCosts           =
as.numeric(0),
                                    D_EarlyTreatmentCosts   =
as.numeric(0),
                                    D_LateTreatmentCosts    =
as.numeric(D_NoScreening_Costs),
                                    N_CancerDeaths          =
as.numeric(N_CancerDeaths),
                                    N_Overdiagnosed         =
as.numeric(N_Overdiagnosed),
                                    N_ClinicalCases         =
as.numeric(N_ClinicalCases)
                                    )
```

32. Simulate screening strategies with a loop for the number of strategies defined previously.

```
  for (Strategy in c(1:length(strategies))){
```

33. Print the progress of the simulation, showing the current run and the current strategy with the total number of simulations and strategies behind the slash separately.

```
    # To track the progress
    print(paste("Simulation ", CurrentRun, " / ", ncol(Input), " :
Strategy ", Strategy, " / ", length(strategies), sep = ""))
```

34. Retrieve the parameters for defining the simulated screening programme.

```
    # Retrieve the strategy-specific values
    StartAge           <- StartAges[Strategy]
    StopAge            <- StopAges[Strategy]
    IntervalSwitchAge1 <- IntervalSwitchAge1s[Strategy]
    IntervalSwitchAge2 <- IntervalSwitchAge2s[Strategy]
```

```
    IntervalSwitchAge3 <- IntervalSwitchAge3s[Strategy]
    Interval1          <- Interval1s[Strategy]
    Interval2          <- Interval2s[Strategy]
    Interval3          <- Interval3s[Strategy]
    Interval4          <- Interval4s[Strategy]
    TestSwitchAge      <- TestSwitchAges[Strategy]
    ScreenTest1        <- ScreenTest1s[Strategy]
    ScreenTest2        <- ScreenTest2s[Strategy]
```

35. Adjust the screening schedules by rounding the number of screening. If Interval1 equals zero, only one screening is implemented throughout the lifetime and it happens at the screening starting age. If Interval1 is not zero, the screening ages are generated by the following code. First, create two strings to save the intervals, IntervalVector, and the ages changing the screening intervals, ChangeAges. Second, have a loop for the number of intervals. Inside the loop, we update the screening stop age from ChangeAges and use this stop age to find the closest number of screenings. If it can be round up to two numbers, this model chooses the larger number of screenings. Use this number of screenings to calculate the correct stop age and produce the screening ages before this stop age. The maximum of the screening ages is the starting age of the next loop to calculate the number of screenings. Screens records the adjusted screening ages and NumberOfScreens is the total number of screenings.

```
    # Amend the screening schedule
    if (Interval1 == 0){  # One-time screen in the lifetime
      Screens <- StartAge
      NumberOfScreens <- 1
    } else {
        Intervals <- 1  # Determine the number of intervals, and the
default is one interval
        if (!is.na(IntervalSwitchAge1)){Intervals <- 2}
        if (!is.na(IntervalSwitchAge2)){Intervals <- 3}
        if (!is.na(IntervalSwitchAge3)){Intervals <- 4}

        IntervalVector <- c(Interval1, Interval2, Interval3,
Interval4)  # Create a vector of the screening intervals
        ChangeAges <- sort(na.omit(c(StopAge, IntervalSwitchAge1,
IntervalSwitchAge2, IntervalSwitchAge3)))  # Create a vector of ages
at which screening changes
        Screens <- NULL  # Create an empty vector to bind the sequence
of screens
        for (Interval in 1:Intervals ){  # Loops through the Intervals
          StopAge <- ChangeAges[Interval]  # Redefine the StopAge as
the ChangeAge in this interval
          NumberOfScreens <- (StopAge - StartAge) /
IntervalVector[Interval] + 1  # Find the number of screens
          NumberOfScreens <- ifelse(NumberOfScreens %% 1 == 0.5,
ceiling(NumberOfScreens), round(NumberOfScreens))  # Rounding the
```

```
number of screens
          StopAge <- StartAge + (NumberOfScreens - 1) *
IntervalVector[Interval]  # Redefine the actual StopAge following
rounding
          Screens <- c(Screens, seq(StartAge, StopAge,
IntervalVector[Interval]))  # Create the sequence of screens from the
start and stop ages
          StartAge <- max(Screens)  # Update the final screen age as
the Start Age for cases in which the interval changes
        }
        Screens <- unique(Screens)  # Remove duplicates from the
screen schedule
        NumberOfScreens <- length(Screens)  # Update the number of
screens
    }
```

36. Generate an array ScreenCounts to record the outcomes of screening. The first three columns record the background information of screening, including the screening number for tracking which round of screening is implemented, the applied screening modality, and the screening age. Before the test switch age, TestSwitchAge, the ScreenTest1 is the applied screening modality. After the test switch age, the screening modality, ScreenTest2, is used. The rest four columns are the number of individuals receiving the screening, the number of individuals in the preclinical state that can be diagnosed by the screening (although this is unknown in reality), the number of true positives, and the number of false positives.

```
    # Create an array of the screen counts: This is an array recording
the number of screens and consequent results
    ScreenCounts <- array(NA, dim = c(NumberOfScreens, 7))  # Create
an array from the screen schedule
    colnames(ScreenCounts) <- c("ScreenNumber", "TestApplied",
"ScreenAge", "N_Screens", "RealPostives", "TruePositives",
"FalsePositives")  # Name the columns
    ScreenCounts[, "ScreenNumber"] <- c(1:NumberOfScreens)  # Write in
a list equivalent to the number of screens and the per round screening
age
    ScreenCounts[, "ScreenAge"] <- Screens
    ScreenCounts[, "TestApplied"] <- ScreenTest1  # Insert the
appropriate test number
    ScreenCounts[which(ScreenCounts[, "ScreenAge"] >= TestSwitchAge),
"TestApplied"] <- ScreenTest2
```

37. Prepare for screening strategy simulation. Create an outcome table, ScreenedOutcomes, to record individuals' disease history. Create two variables to record the clinical and cured cases' IDs.

```
    # Now run screening
    ScreenedOutcomes <- Outcomes  # Create the screen-adjusted
outcomes from the natural history outcomes
    ScreenedClinicalCases <- NULL  # Create an empty vector to record
clinical cases
    ScreenedCuredCases <- NULL  # Create an empty vector to record
cured cases
```

38. Simulate the screening intervention. Create a loop for the number of screens. Within each loop, reset the sample seed to have a constant chance of being diagnosed for the same age across the screening programmes. In this example, we have **set.seed** refer to the table of random numbers of the screening age plus 2020. The user can modify 2020 to any number or delete it, as long as it is fixed across the screening strategies. The reason for choosing the screening age for **set.seed** is that we expect the screening implemented at a specific age will have the same performance across the screening programmes. Undiagnosed individuals have the same chance detected by the screening if the screening implemented at the same age. This is to minimise the statistical noise from the sampling. This issue can also be solved by increasing the simulation sample size.

After referring to the intended table of random numbers, produce two sets of sample numbers for all the individuals, separately for test sensitivity and test specificity. Retrieve relevant parameter values, including screening age, test sensitivity, and test specificity.

Identify individuals who are eligible for screening. First, these individuals must be alive at the moment of implementing screening. These individuals can be found by comparing their all-cause deaths and the screening age. Second, these individuals cannot be already diagnosed at the moment of implementing screening, so compare their clinical ages and the screening age. Notably, individuals who never develop the disease can be screened as well. For those who do not have clinical ages, their results of comparison to the screening age are NA. We modify these NA to *TRUE*, so these people can receive the screening. Only when individuals meet the first and the second conditions, they are qualified for receiving screening.

Recognise all the individuals can be screened for positives and negatives. Individuals who can be screened for positives, AllPositives, are the individuals qualified for screening and in the preclinical state. These individuals have entered the preclinical state at the moment of screening. Compare their ages of entering preclinical state and the screening age and make sure individuals who never enter the preclinical state do not meet these criteria. Individuals who can be screened for negatives, AllNegatives, are the individuals who are qualified for screening but not in the preclinical state.

Distinguish the true positives and false positives by comparing the sample numbers and the test sensitivity and specificity. Test sensitivity is the percentage of correctly identifying patients with the disease. Conversely, test specificity is the percentage of correctly identifying patients without the disease. Accordingly, we apply two different cohorts, AllPositives and AllNegatives, to separately find the true positives and false positives.

Record the results back to table ScreenCounts, including the number of individuals receiving the screening, the number of people who can be screened positive, the number of true positives, and the number of false positives.

Observe who can be cured by the treatment. We refer CureSeed as the random numbers which are produced when we simulate the no-screening scenario, so individuals have compatible results. This can avoid the situation that an individual fails the treatment in the screening programme and is alive in the no-screening scenario despite a higher treatment success chance with the screening. The impact of this issue can also be minimized by simulating a larger cohort.

Finally, update the individuals' natural history with the screening intervention. For the true-positives, update their age of being diagnosed (entering the clinical state). For those cured by treatment, update their all-cause deaths with their other-cause deaths.

Record the clinical cases' IDs to `ScreenedClinicalCases` and cured cases' IDs to `ScreenedCuredCases`.

```
    for (ScreenNumber in 1:nrow(ScreenCounts)){
      set.seed(ScreenCounts[ScreenNumber, "ScreenAge"] + 2020)  #
Fixed sample seed for the same screening age across screening
programmes
      ScreenSnSeed <- runif(SampleSize)  # Save random numbers for the
test sensitivity
      ScreenSpSeed <- runif(SampleSize)  # Save random numbers for the
test specificity
      # Define the screen age
      ScreenAge <- ScreenCounts[ScreenNumber, "ScreenAge"]
      # Retrieve the test sensitivity and specificity
      ScreenSn <- Input[paste("TestSensitivity",
ScreenCounts[ScreenNumber, "TestApplied"], sep = ""), CurrentRun]
      ScreenSp <- Input[paste("TestSpecificity",
ScreenCounts[ScreenNumber, "TestApplied"], sep = ""), CurrentRun]

      Alive <- ScreenedOutcomes[, "AllCauseDeath"] >= ScreenAge
      NotDiagnosed <- ScreenedOutcomes[, "Clinical"] >= ScreenAge
      NotDiagnosed[is.na(NotDiagnosed)] <- TRUE  # People who never
develop the disease are also not diagnosed
      ScreenEligible <- Alive * NotDiagnosed  # Only when both
conditions (alive and not diagnosed) meet

      # Identify those in the preclinical stage at the screen age
      Preclinical <- ScreenedOutcomes[, "Preclinical"] <= ScreenAge
      Preclinical[is.na(Preclinical)] <- FALSE  # People who never
develop the disease are not in the preclinical stage
      AllPositives <- which((Preclinical * ScreenEligible) == 1)  #
Save patient numbers that can be screened for positive
      # Identify the negatives as the complement of the positives from
within the ScreenEligible set
      AllNegatives <- which(ScreenEligible ==
1)[!(which(ScreenEligible == 1) %in% AllPositives)]

      # Find the true positives by sampling without replacement over
all positives in proportion to the test sensitivity
      TruePositives <- AllPositives[ScreenSnSeed[AllPositives] <=
ScreenSn]
      # Find the false positives by sampling without replacement over
the negatives in proportion to the test specificity
      FalsePositives <- AllNegatives[ScreenSpSeed[AllNegatives] >=
ScreenSp]
      # It's useful to have a per screening round count of the number
of positives, all positives, true positives and false positives
      ScreenCounts[ScreenNumber, c("N_Screens", "RealPostives",
"TruePositives", "FalsePositives")] <-
cbind(length(which(ScreenEligible == 1)),
                          length(AllPositives),
                          length(TruePositives),
                          length(FalsePositives)
                          )
      # Censor these successfully treated individuals
      ScreenedCured <- TruePositives[which(CureSeed[TruePositives] <=
Input["PreClinicalProbability", CurrentRun])]
```

```
      # Now update the screen-adjusted outcomes
      ScreenedOutcomes[TruePositives, "Clinical"] <- ScreenAge  #
Update the age of entering clinical state
      ScreenedOutcomes[ScreenedCured, "AllCauseDeath"] <-
ScreenedOutcomes[ScreenedCured, "OtherCauseDeath"]  # Update all-cause
death
      # Record clinical cases
      ScreenedClinicalCases <- c(ScreenedClinicalCases, TruePositives)
# Save patient numbers that are diagnosed by screening
      ScreenedCuredCases <- c(ScreenedCuredCases, ScreenedCured)  #
Save patient numbers that are cured by screening
    }
```

39. Save the intermediate outcomes of screening performance, `ScreenCounts`, to the following folder.

```
    # Record intermediate outcomes
    write.table(ScreenCounts,
paste("OutputFiles/LatestResults/IntermediateOutcomtes/ScreenCounts/",
colnames(Input)[CurrentRun], "_", strategies[Strategy], ".txt", sep =
""), row.names = F, col.names = T, sep = '\t')
```

40. Calculate the number of over-diagnosis, clinical cases, and cancer deaths for the intermediate outcomes. The over-diagnoses are the individuals dying from other-cause before they are diagnosed in their natural disease history, but unnecessarily diagnosed by the screening intervention. The clinical cases are the patients that are diagnosed in their disease history or by the screening programmes. `NotScreenedClinicalCases` are the patients diagnosed in their disease history but not the screening, despite the implementation of screening programmes. `N_CancerDeaths` is the number of patients whose cause-specific death is equal to its all-cause death.

```
    N_Overdiagnosed <- sum(!ScreenedClinicalCases %in% ClinicalCases)
# Patients are not diagnosed in their natural history but diagnosed
with screening
    N_ClinicalCases <- length(ClinicalCases) +
sum(!ScreenedClinicalCases %in% ClinicalCases)  # The number of
patients diagnosed in natural history and screening
    NotScreenedClinicalCases <- ClinicalCases[!ClinicalCases %in%
ScreenedClinicalCases]  # Patient numbers who are diagnosed without
screening
    NotScreenedCuredCases <-
NotScreenedClinicalCases[which(CureSeed[NotScreenedClinicalCases] <=
Input["ClinicalProbability", CurrentRun])]  # Receive treatment with
```

```
clinical probability
    N_CancerDeaths <- length(which(ScreenedOutcomes[, "AllCauseDeath"]
== ScreenedOutcomes[, "CauseSpecificDeath"]))  # The number of cases
that die at the age of disease death
```

41. Apply the self-defined function AgeDistribution to the Outcomes table which records individuals' disease history. This function is defined in step 29. Then, save the results to the following folder.

```
    # Record age distribution of each health state
    IntermediateOutcomes_AgeDist <- apply(ScreenedOutcomes[,
!(colnames(ScreenedOutcomes) %in% "PersonNumber")], 2,
AgeDistribution)
    write.table(IntermediateOutcomes_AgeDist,
paste("OutputFiles/LatestResults/IntermediateOutcomtes/AgeDistribution
/", colnames(Input)[CurrentRun], "_", strategies[Strategy], ".txt",
sep = ""), row.names = F, col.names = T, sep = '\t')
```

42. Prepare for calculating quality-of-life adjusted outcomes, similar to step 21 for the no-screening scenario. First, we save the discount and undiscounted life-years from the ScreenedOutcomes. Apply the PresentValue function defined in step 20 to discount the life-years.

```
    # Assess effects
    UD_QALYS <- ScreenedOutcomes[, "AllCauseDeath"]
    D_QALYS <- D_LYS <- sapply(UD_QALYS, PresentValue)  # Apply the
discounting function that we defined above
```

43. Update the disease history of sick individuals, SickOutcome. We assume the quality-of-life (utility weight) only changes when individuals enter the clinical states in natural history. Early diagnosis does not influence the quality-of-life, while clinically having symptoms reduces the quality-of-life. The new SickOutcome has the natural disease history for all health states, except for all-cause deaths which will be replaced by the screening outcomes.

```
    # Quality-of-life only changes when entering another health state
in their natural history
    SickOutcome <- cbind(SickOutcome[ , - length(StagesInOrder)],
ScreenedOutcomes[Sick, "AllCauseDeath"])  # Update the age of death in
the natural history of all the sick people
```

44. Apply the `PresentValue` function defined in step 20 to discount the ages of health states and apply the `Accumulated_QALYS` function defined in step 22 to calculate the QALYs for the sick. For individuals who never develop the disease, adjust their QALYs with the utility weight of the healthy state.

```
    Dis_SickOutcome <- apply(SickOutcome, MARGIN = c(1, 2),
PresentValue)  # Apply the discounting function

    UD_QALYS[DiseaseFree] <- UD_QALYS[DiseaseFree] * Utility[1]  #
Quality-of-life adjusted but not discounted
    UD_QALYS[Sick] <- apply(SickOutcome, 1, Accumulated_QALYS)
    D_QALYS[DiseaseFree] <- D_QALYS[DiseaseFree] * Utility[1]  #
Quality-of-life adjusted and discounted
    D_QALYS[Sick] <- apply(Dis_SickOutcome, 1, Accumulated_QALYS)
```

45. Calculate discounted and undiscounted QALYs loss due to screening and triage. Create a variable `DiscountFactor_Screen` to record the discounting factors corresponding to the screening ages. These are applied to discount the QALYs loss from screening and triage. The total QALYs loss due to screening is the number of screenings multiply the dis-utility of one screening test. Notably, the test dis-utility might be different if the screening strategy changes the screening modality in the middle of the programme, so we refer to the modality specified in the `ScreenCounts`. The total triage service is the number of true-positives and false-positives multiply the dis-utility of one triage.

```
    # Calculate utility loss due to screening test
    DiscountFactor_Screen <- DF_Effects[ScreenCounts[, "ScreenAge"]]
    UD_ScreenHarm <- sum(ScreenCounts[,"N_Screens"] *
Input[paste("TestDisutility", ScreenCounts[,"TestApplied"], sep=""),
CurrentRun])
    D_ScreenHarm <- sum(ScreenCounts[,"N_Screens"] *
Input[paste("TestDisutility", ScreenCounts[,"TestApplied"], sep=""),
CurrentRun] * DiscountFactor_Screen)  # Discounted
    UD_TriageHarm <- sum((ScreenCounts[, "TruePositives"] +
ScreenCounts[, "FalsePositives"]) * Input["DisutilityTriage",
CurrentRun])  # All the test positives are triaged
    D_TriageHarm <- sum((ScreenCounts[, "TruePositives"] +
```

```
ScreenCounts[, "FalsePositives"]) * Input["DisutilityTriage",
CurrentRun] * DiscountFactor_Screen)
```

46. Calculate discounted and undiscounted QALYs loss due to treatment. A perfect triage is assumed, so only the true-positives receive the treatments. The treatment is assumed to be carried out at the end of the year of diagnosis. For patients who are diagnosed in their natural history, round up their ages of being diagnosed in the ScreenedOutcomes and find the corresponding discounting factors. For patients who are diagnosed in the screening programmes, apply the corresponding discounting factor of screening ages, DiscountFactor_Screen defined in step 45.

```
    # Calculate utility loss due to treatment
    ClinicalOnsetAges <-
ceiling(ScreenedOutcomes[NotScreenedClinicalCases, "Clinical"])  #
Identify the clinical ages and round up to integers for discounting
    UD_TreatmentHarm <- sum(ScreenCounts[,"TruePositives"] *
Input["DisutilityTrt", CurrentRun]) +   # Diagnosed with screening
                        length(NotScreenedClinicalCases) *
Input["DisutilityTrt", CurrentRun]  # Diagnosed without screening
    D_TreatmentHarm <- sum(ScreenCounts[,"TruePositives"] *
Input["DisutilityTrt", CurrentRun] * DiscountFactor_Screen) +   #
Receive treatment at the age of screening
                       sum(Input["DisutilityTrt", CurrentRun] *
DF_Effects[ClinicalOnsetAges])  # Receive treatment when entering
clinical state in natural history
```

47. Sum up the life-years and QALYs of each individual, including discounted and undiscounted. The calculation of QALYs for the screening intervention needs to deduct the dis-utility from screening, triage, and treatment.

```
    # Calculate the discounted life-years without screening
    UD_Strategy_LYG <- sum(ScreenedOutcomes[, "AllCauseDeath"])
    D_Strategy_LYG <- sum(D_LYS)
    UD_Strategy_QALY <- sum(UD_QALYS) - UD_ScreenHarm - UD_TriageHarm
- UD_TreatmentHarm
    D_Strategy_QALY <- sum(D_QALYS) - D_ScreenHarm - D_TriageHarm -
D_TreatmentHarm
```

48. Estimate the discounted and undiscounted treatment costs of patients diagnosed clinically. In the screening programmes, some patients are diagnosed with screening and

some patients are diagnosed with symptoms when they enter the clinical state in their natural disease history. In this step, we calculate the treatment costs for patients clinically presented by the number of NotScreenedClinicalCases multiply the unit cost of the treatment for the systematic disease. For discounted late treatment costs, apply the discounting factors that are based on the discounting rate of costs and correspond to the ages entering clinical state. Also, update the variable DiscountFactor_Screen with the discounting rate of costs.

```
    # Assess costs
    # Count the late treatment costs
    DiscountFactor_Screen <- DF_Costs[ScreenCounts[, "ScreenAge"]]
    UD_LateTreatmentCosts <- length(NotScreenedClinicalCases) *
LateTreatmentCost
    D_LateTreatmentCosts <- sum(DF_Costs[ClinicalOnsetAges] *
LateTreatmentCost)  # Find the discounted treatment costs
```

49. Estimate the undiscounted costs of screening, triage, and treatment for the screening-detected. For undiscounted screening costs, add up the number of individuals receiving the screening and times to the unit screening cost. The unit triage cost applies to all the true-positives and the false-positives. The treatement costs for the screening-detected is only for those are true-postives.

```
    # First calculate the undiscounted amounts: the costs of primary
screening, triage and treating preclinical disease under screening
    UD_ScreenCosts <- sum(ScreenCounts[, "N_Screens"] * ScreenCost)  #
The primary test cost applies to all screens
    UD_TriageCosts <- sum((ScreenCounts[, "TruePositives"] +
ScreenCounts[, "FalsePositives"]) * TriageCost)  # The cost of triage
applies to all screen true and false positives
    UD_EarlyTreatmentCosts <- sum(ScreenCounts[, "TruePositives"] *
EarlyTreatmentCost)  # The early treatment costs applies to all true
positives (assuming a 100% specific triage test)
```

50. Estimate the discounted costs of screening, triage, and treatment for the screening-detected. Create an array ScreenCostArray to save the screening ages, discounted screening costs, discounted triage costs, and discounted treatment costs. Use the DiscountFactor_Screen defined in step 48 to discount the cost estimates for the screen-detected.

```
    # Now calculate the discounted values
    ScreenCostArray <- array(NA, dim = c(NumberOfScreens, 4))  #
Create a blank array for the discounted costs which can also be saved
as intermediate outcomes
    colnames(ScreenCostArray) <- c("Age", "D_Screen", "D_Triage",
"D_EarlyTreatment")  # Name the columns in this array
    ScreenCostArray[, "Age"] <- ScreenCounts[, "ScreenAge"]  # Apply
the appropriate discount factors according to the screen ages
    ScreenCostArray[, "D_Screen"] <- DiscountFactor_Screen *
ScreenCounts[, "N_Screens"] * ScreenCost  # Multiply the screen,
triage and early treatment numbers by the discount factor
    ScreenCostArray[, "D_Triage"] <- DiscountFactor_Screen *
(ScreenCounts[, "TruePositives"] + ScreenCounts[, "FalsePositives"]) *
TriageCost
    ScreenCostArray[, "D_EarlyTreatment"] <- DiscountFactor_Screen *
ScreenCounts[, "TruePositives"] * EarlyTreatmentCost
```

51. Calculate the undiscounted and discounted total costs. Sum up the costs of screening, triage, and treatment for both symptomatic disease and screening-detected.

```
    # Sum up the undiscounted and discounted costs
    UD_TotalCosts <- UD_ScreenCosts + UD_TriageCosts +
UD_EarlyTreatmentCosts + UD_LateTreatmentCosts
    D_TotalCosts <- sum(ScreenCostArray[, c("D_Screen", "D_Triage",
"D_EarlyTreatment")]) + D_LateTreatmentCosts
```

52. Update the `IntermediateOutcomes` with the results for this specific screening strategy.

```
    # Record intermediate outcomes and the first column is the name of
the strategy
    IntermediateOutcomes[Strategy + 1, - 1] <- c(UD_ScreenCosts
= UD_ScreenCosts,
                                                 UD_TriageCosts
= UD_TriageCosts,
                                                 UD_EarlyTreatmentCosts
= UD_EarlyTreatmentCosts,
                                                 UD_LateTreatmentCosts
= UD_LateTreatmentCosts,
                                                 D_ScreenCosts
= sum(ScreenCostArray[, "D_Screen"]),
                                                 D_TriageCosts
= sum(ScreenCostArray[, "D_Triage"]),
                                                 D_EarlyTreatmentCosts
= sum(ScreenCostArray[, "D_EarlyTreatment"]),
                                                 D_LateTreatmentCosts
= D_LateTreatmentCosts,
                                                 N_CancerDeaths,
```

```
                                                    N_Overdiagnosed,
                                                    N_ClinicalCases
                                                    )
```

53. Update the CostEffectivenessResults with the simulation outcomes for the specific screening strategy. End the loop of the simulation if it is the last screening strategy.

```
    # Sum up the undiscounted and discounted costs
    CostEffectivenessResults[Strategy + 1, c("UD_EffectsL",
"UD_EffectsQ", "UD_Costs", "D_EffectsL", "D_EffectsQ", "D_Costs")] <-
c(UD_Strategy_LYG,
                                                    UD_Strategy_QALY,
                                                    UD_TotalCosts,
                                                    D_Strategy_LYG,
                                                    D_Strategy_QALY,
                                                    D_TotalCosts
                                                    )
  }  # Close the loop over all the strategies
```

54. Create a self-defined function SDfinder to search for the simple dominance among the screening strategies. Strategies are simple dominated (SD) if they, compared to any other strategy, have lower effectiveness but higher costs or higher costs but lower effectiveness. Strategies lablled with NA mean they are not SD.

```
  # Mark those subject to simple dominance as SD
  SDfinder <- function(x, output){  # This simply compares each
strategy to all others to find cases of simple dominance
    if (sum((as.numeric(x[1]) <= AllEffects) * (as.numeric(x[2]) >
AllCosts)) >= 1){
      return("SD")
    }
    if (sum((as.numeric(x[1]) < AllEffects) * (as.numeric(x[2]) >=
AllCosts)) >= 1){
      return("SD")
    }
    return("NA")
  }
```

55. Employ the function SDfinder defined in step 54 to identify the strategies that are SD with the discounted and undiscounted results, including LYs and QALYs.

```
  AllCosts <- as.numeric(CostEffectivenessResults[, "D_Costs"])
  AllEffects <- as.numeric(CostEffectivenessResults[, "D_EffectsL"])
  CostEffectivenessResults[, "D_ICER_L"] <-
apply(CostEffectivenessResults[, c('D_EffectsL', 'D_Costs')], 1,
SDfinder)  # Discounted LYs gained
  AllEffects <- as.numeric(CostEffectivenessResults[, "D_EffectsQ"])
  CostEffectivenessResults[, "D_ICER_Q"] <-
apply(CostEffectivenessResults[, c('D_EffectsQ', 'D_Costs')], 1,
SDfinder)  # Discounted QALYs gained
  AllCosts <- as.numeric(CostEffectivenessResults[, "UD_Costs"])
  AllEffects <- as.numeric(CostEffectivenessResults[, "UD_EffectsL"])
  CostEffectivenessResults[, "UD_ICER_L"] <-
apply(CostEffectivenessResults[, c('UD_EffectsL', 'UD_Costs')], 1,
SDfinder)  # Undiscounted LYs gained
  AllEffects <- as.numeric(CostEffectivenessResults[, "UD_EffectsQ"])
  CostEffectivenessResults[, "UD_ICER_Q"] <-
apply(CostEffectivenessResults[, c('UD_EffectsQ', 'UD_Costs')], 1,
SDfinder)  # Undiscounted QALYs gained
```

56. Create a self-defined function EDfinder to identify the strategies which are extended dominated (ED). First, create a BoundarySet that excludes the SD strategies. In the case of BoundarySet only have a single strategy or all the strategies with the same cost and effecitveness, only one cost-effectiveness estimate on the efficient frontier. This cost-effectiveness estimate should be the reference case without numeric ICER.

In other circumstances, order the strategies from the least effective to the highest. The least effective strategy is the reference, the first strategy on the efficient frontier. If there is only one strategy or if all the strategies have the same effectiveness and costs, only one reference case lies on the efficient frontier without any ICER estimations. If there is more than one strategy having different outcomes, calculate the cost-effectiveness ratio compared to the reference case for all the strategies having higher costs and effectiveness than the reference case. Find the strategy with the lowest ratio as the next strategy on the efficient frontier. Compared to this strategy, the strategies having lower effectiveness are extended dominated (ED). Then, exclude these ED strategies and recalculate the cost-effectiveness ratio compared to the new strategy identified on the efficient frontier. This loop continues until the strategy with the largest effectiveness is identified. This function returns BoundarySet, the strategies are ED or on the efficient frontier.

```
  # Find cost-effective screening strategies on the efficient frontier
  EDfinder <- function(CEresults, Effects, Costs, ICER, output){
```

```
    # Define the boundary set as the non strictly-dominated strategies
    BoundarySet <- CEresults[setdiff(seq(1:nrow(CEresults)),
which(CEresults[, ICER] == "SD")),]
    # Order the boundary set in terms of effectiveness
    BoundarySet <- BoundarySet[order(BoundarySet[, Effects]),]
    # The first efficient strategy is the strategy with the least
effectiveness
    EfficientSet <- 1

    if (nrow(BoundarySet) == 1 | (length(unique(BoundarySet[,
Effects])) == 1 & length(unique(BoundarySet[, Costs])) == 1)){  # Only
one strategy or strategies with the same outcomes
      BoundarySet[, ICER] <- "reference"  # This strategy is the only
strategy on the efficient frontier, so without ICER
    } else {
        # Set the ICERs as numeric
        BoundarySet[, ICER] <- as.numeric(BoundarySet[, ICER])
        # Put a the break condition at the beginning that stops the
routine if the end of the set is reached
        if (max(EfficientSet) < nrow(BoundarySet)){
          repeat{
            # Update the ICER calculation relative to the new
reference
            for (i in ((max(EfficientSet) + 1): nrow(BoundarySet))){
              # Calculate the column of ACERs relative to the least-
effective non simply dominated strategy
              BoundarySet[i, ICER] <- round((BoundarySet[i, Costs] -
BoundarySet[max(EfficientSet), Costs])/
                                              (BoundarySet[i, Effects] -
BoundarySet[max(EfficientSet), Effects]), 3)
            }
            # Find the lowest ICER
            j <- max(EfficientSet) +
which(BoundarySet[(max(EfficientSet) + 1):nrow(BoundarySet), ICER] ==

min(BoundarySet[(max(EfficientSet) + 1):nrow(BoundarySet), ICER],
na.rm=TRUE))
            # Update what forms the efficient set
            EfficientSet <- c(EfficientSet, j)

            if (max(EfficientSet) == nrow(BoundarySet)){break}
          }
        }
        BoundarySet[-EfficientSet, ICER] <- "ED"
        # Set the ICER of the least effective strategy in the
efficient set to "reference"
        BoundarySet[1, ICER] <- "reference"
      }
    return(BoundarySet)
  }
```

57. Employ the function `EDfinder` defined in step 56 to identify the strategies that are ED with the discounted and undiscounted results, including LYs and QALYs.

```
  BoundarySet_L <- EDfinder(CostEffectivenessResults, "D_EffectsL",
"D_Costs", "D_ICER_L")
  BoundarySet_Q <- EDfinder(CostEffectivenessResults, "D_EffectsQ",
"D_Costs", "D_ICER_Q")
  BoundarySet_UL <- EDfinder(CostEffectivenessResults, "UD_EffectsL",
"UD_Costs", "UD_ICER_L")
  BoundarySet_UQ <- EDfinder(CostEffectivenessResults, "UD_EffectsQ",
"UD_Costs", "UD_ICER_Q")
```

58. Write the calculated ICERs and ED to the main outcome table CostEffectivenessResults.

```
  # Write in the ICERs of the efficient strategies
  CostEffectivenessResults[match(BoundarySet_L[, "StrategyName"],
CostEffectivenessResults[, "StrategyName"]), "D_ICER_L"] <-
BoundarySet_L[, "D_ICER_L"]
  CostEffectivenessResults[match(BoundarySet_Q[, "StrategyName"],
CostEffectivenessResults[, "StrategyName"]), "D_ICER_Q"] <-
BoundarySet_Q[, "D_ICER_Q"]
  CostEffectivenessResults[match(BoundarySet_UL[, "StrategyName"],
CostEffectivenessResults[, "StrategyName"]), "UD_ICER_L"] <-
BoundarySet_UL[, "UD_ICER_L"]
  CostEffectivenessResults[match(BoundarySet_UQ[, "StrategyName"],
CostEffectivenessResults[, "StrategyName"]), "UD_ICER_Q"] <-
BoundarySet_UQ[, "UD_ICER_Q"]
```

59. Update the tables BoundarySet_L, BoundarySet_Q, BoundarySet_UL, and BoundarySet_UQ to record the boundary sets for discounted LYs, discounted QALYs, undiscounted LYs, and undiscounted QALYs. For each table, exclude ED strategies and only leave strategies that are on the efficient frontier.

```
  BoundarySet_L <- BoundarySet_L[which(!BoundarySet_L[, "D_ICER_L"]
%in% "ED"), ]  # Update the BoundarySet by removing strategies with
"ED"
  BoundarySet_Q <- BoundarySet_Q[which(!BoundarySet_Q[, "D_ICER_Q"]
%in% "ED"), ]
  BoundarySet_UL <- BoundarySet_UL[which(!BoundarySet_UL[,
"UD_ICER_L"] %in% "ED"), ]
  BoundarySet_UQ <- BoundarySet_UQ[which(!BoundarySet_UQ[,
"UD_ICER_Q"] %in% "ED"), ]
```

60. Save the main results, CostEffectivenessResults, and intermediate outcomes, IntermediateOutcomes, for all the screening strategies to the assigned folder. Also, the

strategies on the efficient frontier are also saved in independent files for different outcomes, discounted or undiscounted, QALYs or LYs.

```
  # Return the results for all values and the efficient set within the
R session
  # Save the overall results and efficient frontier results
  write.table(IntermediateOutcomes,
paste("OutputFiles/LatestResults/IntermediateOutcomtes/CEbreakdown/",
colnames(Input)[CurrentRun], ".txt", sep = ""),
              row.names = F, col.names = T, sep = '\t')
  write.table(CostEffectivenessResults,
paste("OutputFiles/LatestResults/MainCEresults/OutputTable_",
colnames(Input)[CurrentRun], ".txt", sep=""),
              row.names = F, col.names = T, sep = '\t')
  write.table(BoundarySet_L,
paste("OutputFiles/LatestResults/MainCEresults/BoundarySet/BoundarySet
_L_", colnames(Input)[CurrentRun], ".txt", sep=""),
              row.names = F, col.names = T, sep = '\t')
  write.table(BoundarySet_Q,
paste("OutputFiles/LatestResults/MainCEresults/BoundarySet/BoundarySet
_Q_", colnames(Input)[CurrentRun], ".txt", sep=""),
              row.names = F, col.names = T, sep = '\t')
  write.table(BoundarySet_UL,
paste("OutputFiles/LatestResults/MainCEresults/BoundarySet/BoundarySet
_UL_", colnames(Input)[CurrentRun], ".txt", sep=""),
              row.names = F, col.names = T, sep = '\t')
  write.table(BoundarySet_UQ,
paste("OutputFiles/LatestResults/MainCEresults/BoundarySet/BoundarySet
_UQ_", colnames(Input)[CurrentRun], ".txt", sep=""),
              row.names = F, col.names = T, sep = '\t')
```

61. Plot the cost-effectiveness plane only when simulating the base-case scenario and the user intends to have the cost-effectiveness plane.

```
  # Plot cost-effectiveness plane
  if ((colnames(Input)[CurrentRun] == "Input") & (CEplane == TRUE)){
# Only plot when it is the base-case
```

62. Plot the cost-effectiveness plane with the user-defined outcomes, QALYs or LYs.

```
    # Define the outcome measure
    if (MeasureQALYs == TRUE){
      OutcomeMeasure <- "D_EffectsQ"
      ICERMeasure <- "D_ICER_Q"
      BoundarySet <- BoundarySet_Q
```

```
    } else {
        OutcomeMeasure <- "D_EffectsL"
        ICERMeasure <- "D_ICER_L"
        BoundarySet <- BoundarySet_L
      }
```

63. Have all the results reference back to the no-screening strategy which is the first row of the outcome tables, including CostEffectivenessResults and BoundarySet. The no-screening will be the origin of the cost-effectiveness plane. Also, having the average effectiveness and costs per capita is easier to compare the results when simulating different sample sizes.

```
    # Define the costs and effects, reference to the no-screening
    Effects <- (CostEffectivenessResults[, OutcomeMeasure] -
CostEffectivenessResults[1, OutcomeMeasure]) / SampleSize
    Costs <- (CostEffectivenessResults[, "D_Costs"] -
CostEffectivenessResults[1, "D_Costs"]) / SampleSize
    BoundarySet$EfficientEffects <- (BoundarySet[, OutcomeMeasure] -
BoundarySet[1, OutcomeMeasure]) / SampleSize
    BoundarySet$EfficientCosts <- (BoundarySet[, "D_Costs"] -
BoundarySet[1, "D_Costs"]) / SampleSize
```

64. Identify the optimal strategy. If the minimum of the available ICERs is below the decision threshold, the optimal strategy is the one with the maximum ICER just lower than the threshold. If all the strategies' ICERs are above the decision threshold, no ICER is available and the reference is the optimal strategy.

```
    # Identify the highest of the cost-effective ICERs given the
threshold
    ICERs <- as.numeric(BoundarySet[, ICERMeasure])  # All the ICERs
    if (min(ICERs, na.rm = TRUE) <= Threshold){
      OptimalStrategy <- rownames(BoundarySet)[which(ICERs ==
(max(ICERs[ICERs <= Threshold], na.rm=TRUE)))]  # Find the maximum
ICER lower than the threshold
    } else {
        OptimalStrategy <- rownames(BoundarySet)[BoundarySet[,
ICERMeasure] %in% "reference"]  # All the ICERs are higher than the
threshold, so the reference is the optimal strategy
      }
```

65. Look up the cost and effectiveness estimates and corresponding ICER of the optimal strategy.

```
    CostEffectiveICER <- BoundarySet[OptimalStrategy, ICERMeasure]  #
Find the corresponding ICER
    CostEffectiveEffects <- as.numeric(BoundarySet[OptimalStrategy,
"EfficientEffects"])  # Find the corresponding costs and effect
    CostEffectiveCosts <- as.numeric(BoundarySet[OptimalStrategy,
"EfficientCosts"])
```

66. Set up the tick marks for axes and save their ranges. The setup of `TicksEffects` and `TicksCosts` may differ from case to case.

```
    # Set the tick marks: after plotting several times, these numbers
are chosen to best present the axis in our example
    TicksEffects <- pretty(c(round(min(Effects) - (max(Effects) -
min(Effects)) / 20, 5), round(max(Effects) + (max(Effects) -
min(Effects)) / 20, 5)))
    TicksCosts <- pretty(c(0, max(Costs) * 1.02))
    # Set the ranges for both values
    xRange <- range(TicksEffects)
    yRange <- range(TicksCosts)
```

67. Generate the ICER label and axis label for the measurement of effectiveness.

```
    # Create ICER and axis labels
    ICERLabels <- gsub("NA", "", noquote(format(BoundarySet[,
ICERMeasure], big.mark = "")))
    if (MeasureQALYs == TRUE){LabLabel <- "Effects (QALY, per
capita)"} else {LabLabel <- "Effects (LYG, per capita)"}
```

68. Create a plot window of a particular size.

```
    # Set the plot dimension
    windows.options(width = 12, height = 9)
```

69. Set the margin sizes of the plot in the following order: bottom, left, top, and right.

```
    # Set the margin parameters
    par(mar = c(5, 6, 1, 1))
```

70. Plot the cost-effectiveness plane which effectiveness is on the X axis. The user-defined parameter HealthOnXAxis controls which axis effectiveness lies on. If the effectiveness is not on the X axis, step 71 produces the plot with the effectiveness on the opposite axis.

Create a plot with Effects on the X axis, and Costs on the Y axis. Remove X and Y axes tick labels by setting xaxt and yaxt to "n". Delete the X and Y axes labels by modifying xlab and ylab to blank strings. Change axes limits xlim and ylim to xRange and yRange that defined as step 66. Choose the symbol of the points by setting pch to any number between 1 and 25, and the default is the filled circle. Adjust the symbol size with cex and the symbol colour with col.

Place the axes with the function **axis**. The first number is the position of the axis. Argument 1 refers to the X axis and 2 refers to the Y axis. Indicate the tickmarks with at and rotate the axis labels horizontally with las.

Add the titles of the X and Y axes with xlab and ylab. Argument mgp controls the placing of the axis title, axis labels, and axis line. Argument cex.lab modifies the size of the axis label.

Add a line and points with boundary data. Use lwd to control the width and pch to choose the symbol of the points. Add the point of the optimal strategy if the user wants to show the optimal strategy on the plot. Apply cex to control the size of symbols. Use col to manage the colour of the symbols.

Place the labels of ICERs, ICERLabels, to the right of the symbols. First, create a LabelYOffset for specifying the space we want to leave in between. Then, add texts with

the specified position. Argument `pos = 4` means plot the texts to the right. Also, the text size is controlled by `cex`.

```
    if (HealthOnXAxis == TRUE){
      # Plot the plane
      plot(Effects, Costs, yaxt = "n", xaxt = "n", xlab = "", ylab =
"", xlim = xRange, ylim = yRange, pch = 16, cex = 0.5, col = "grey")
      # Add the axes
      axis(1, at = TicksEffects, las = 1)
      axis(2, at = TicksCosts, las = 1)
      # Add in the X and Y axis labels with spacing
      title(ylab = "Costs ($, per capita)", xlab = LabLabel, mgp =
c(3.75,1.75,0), cex.lab = 1.2)
      # Add the efficient frontier
      lines(BoundarySet$EfficientEffects, BoundarySet$EfficientCosts,
lwd = 2)
      points(BoundarySet$EfficientEffects, BoundarySet$EfficientCosts,
pch = 19, lwd = 2)
      # Add the optimal strategy
      if (ShowOptimalStrategy == TRUE){points(CostEffectiveEffects,
CostEffectiveCosts, pch = 19, lwd = 3, cex = 1, col = "forestgreen")}
      # Add ICERs to the frontier
      LabelYOffset <- max(BoundarySet$EfficientCosts) / 10  # Generate
a label offset to place the values at an offset
      if (ShowICERs == TRUE){text(BoundarySet$EfficientEffects,
BoundarySet$EfficientCosts - LabelYOffset, labels = ICERLabels, pos =
4, cex = 0.6)}  # Apply the text
    } else {
```

71. Plot the cost-effectiveness plane which effectiveness is on the Y axis. Similarly to step 70, the only difference is to change all the commands relevant to the axes.

Create a plot with `Costs` on the X axis, `Effects` on the Y axis. Change axes limits `xlim` and `ylim` to `yRange` and `xRange`. Modify the first number of the **`axis`** function to have 1 refer to cost and 2 refer to effectiveness. Adjust the titles of the X and Y axes with `xlab` and `ylab` to have correct labels for costs and effectiveness. All the functions regarding lines, points, and texts have the cost data first specified for plotting on the X axis, followed by the effectiveness for the Y axis. In our example, we create two variables, `LabelYOffset` and `LabelXOffset`, to specify the space between the symbols and the texts. These two variables are flexible and adjustable for best plotting the texts.

More details of arguments can be found in step 70.

```
        plot(Costs, Effects, yaxt = "n", xaxt = "n", xlab = "", ylab =
"", xlim = yRange, ylim = xRange, pch = 16, cex = 0.5, col = "grey")
        # Add the axes
        axis(2, at = TicksEffects, las = 1)
        axis(1, at = TicksCosts, las = 1)
        # Add in the X and Y axis labels with spacing
        title(xlab = "Costs ($, per capita)", ylab = LabLabel, mgp =
c(3.75,1.75,0), cex.lab = 1.2)
        # Add the efficient frontier
        lines(BoundarySet$EfficientCosts,
BoundarySet$EfficientEffects, lwd = 2)
        points(BoundarySet$EfficientCosts,
BoundarySet$EfficientEffects, pch = 19, lwd = 2)
        # Add the optimal strategy
        if (ShowOptimalStrategy == TRUE){points(CostEffectiveCosts,
CostEffectiveEffects, pch = 19, lwd = 3, cex = 1, col =
"forestgreen")}
        # Add ICERs to the frontier
        LabelYOffset <- max(BoundarySet$EfficientEffects) / 18   #
Generate a label offset to place the values at an offset
        LabelXOffset <- max(BoundarySet$EfficientCosts) / 40
        if (ShowICERs == TRUE){text(BoundarySet$EfficientCosts +
LabelXOffset, BoundarySet$EfficientEffects + LabelYOffset, labels =
ICERLabels, pos = 2,cex = 0.6)}  # Apply the text
      }
```

72. Save the plot into the file format of .jpeg as default. The width and the height of the diagram can be adjusted. Have higher resolution, `res`, to ensure the clarity of the diagram.

Save the plot into .pdf or .png if needed.

```
    # Save the graph as a .jpeg as a matter of course
    SaveCEplane <- recordPlot()
    jpeg("OutputFiles/Figures/Figure.jpeg", width = 4000, height =
2400, res = 300)
    replayPlot(SaveCEplane)
    dev.off()

    # Save the graph as a PDF or png if that option is ticked
    if (SaveGraphPDF == TRUE){
      pdf("OutputFiles/Figures/Figure.pdf", width = 15, height = 9)
      replayPlot(SaveCEplane)
      dev.off()
    }
    if (SaveGraphPNG == TRUE){
      png("OutputFiles/Figures/Figure.png", width = 4000, height =
2400, res = 300)
      replayPlot(SaveCEplane)
      dev.off()
    }
  }
}
```

73. Print the simulation time by using the current time minus the variable time we saved in the first step of this simulation.

```
Sys.time() - time  # The execution time
```