Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Modelling the Cost-Effectiveness of Cancer Prevention: Reframing Models to Better Match Policy Questions

James F. O'Mahony, MA.

Doctorate of Philosophy

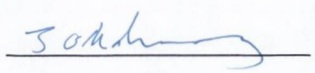
University of Dublin, Trinity College

Supervisors: Prof. Charles Normand & Dr Joost van Rosmalen

Submitted to the University of Dublin, Trinity College,

April 2013.

I, James O'Mahony, declare that this thesis has not been submitted as an exercise for a degree at this or any other University. The work contained in this manuscript includes published and unpublished work undertaken in collaborations with others, as duly acknowledged within the thesis. I agree that the Library may lend or copy the thesis upon request.



SUMMARY

This thesis comprises six papers concerning methodological issues in cost-effectiveness analysis (CEA) of healthcare interventions. The unifying theme to these studies is the reframing of CEAs to better correspond to the policy questions they are to inform.

The first two papers concern the differential discounting of costs and effects in CEA. Differential discounting is the application of a different discount rate to costs and health effects, as distinct from the more usual practice of applying a common discount rate to both. Differential discounting has been advocated as a means of making CEA models more representative of reality, as it accounts for anticipated growth in the value of health over time. What has not been anticipated by advocates of differential discounting are its implications for models featuring multiple future cohorts that start an intervention after the discount year. This leads to the systematic improvement of cost-effectiveness ratios when future cohorts are added to a CEA model. The first paper describes this problem and explains why it leads to difficulties when comparing cost-effectiveness estimates between studies. It uses a CEA of vaccination against the human papillomavirus (HPV) to illustrate the effect of adding multiple future cohorts to an analysis.

The second paper proposes a solution to the problem identified in the first paper. It shows how consideration of a hypothetical intervention that is only marginally cost-effective relative to the current cost-effectiveness threshold can be used to correct cost-effectiveness estimates from models with multiple future cohorts. The application of this method is demonstrated using the example of the results of a previously published CEA of HPV vaccination that features differential discounting and 100 future cohorts.

While many models simulate just one cohort of patients, some use multiple cohorts. Paper 3 describes a feature of such multi-cohort models that has been applied in number of CEA, but not previously described or assessed in the literature. It is the practice of assuming the intervention in question ceases for all cohorts at a given point in time in the future, irrespective of whether or not the recipient cohorts have completed the intervention or not. This is described as the implementation time horizon. When applied in models of interventions that take many years to complete, such as cancer screening programmes, the implementation horizon means that many cohorts will cease the intervention prematurely in the model. Clearly such

premature cessation will be unrepresentative of actual implementation and it raises concerns that CEA models featuring implementation horizons may produce biased cost-effectiveness estimates relative to models in which all cohorts complete their intervention. Paper 3 assesses the implementation horizon using a model used in a recently published CEA of colorectal screening in the Netherlands that assumed all screening stopped 30 years after its introduction. The analysis finds that although the model yields different results when the implementation horizon is relaxed, the differences are minor and would not lead to any meaningful error in the choice of an optimal screening strategy.

Paper 4 addresses another aspect of multi-cohort models. It considers the aggregation of cost-effectiveness estimates over multiple cohorts. While the existing literature on the structure of CEA models of interventions that affect multiple cohorts contemporaneously suggests all cohorts should be modelled together, Paper 4 argues that if cohorts differ meaningfully in their cost-effectiveness, then they should not be aggregated together. The argument is illustrated with a CEA of switching from conventional cytology-based cervical screening to a HPV test.

Paper 5 addresses the modelling of high-risk subgroups in the case of breast cancer screening. It notes that most CEAs are usually conducted for average risk groups and so do not provide useful policy recommendations for the screening of high risk subgroups. Paper 5 describes mathematical model that can be used to adjust optimal screening intervals for an increase in cancer incidence above the average risk rate. The model is verified and validated against a microsimulation model of breast cancer screening.

The final paper in the thesis considers what cervical screening strategies are relevant comparators when estimating the cost-effectiveness of a HPV vaccination. It shows that many existing CEAs of HPV vaccination only consider the current intensity of cervical screening, but that alternative screening intensities are relevant comparators and should be considered. It illustrates the issue by expanding a previously published CEA model of HPV screening in the Netherlands to consider multiple screening intervals and start and stop ages. It shows that considering alternative screening strategies results in a higher incremental cost-effectiveness ratio of vaccination than is found when the current screening strategy is assumed to be the comparator.

ACKNOWLEDGEMENTS

Acknowledgement and thanks go to all the co-authors of work contained in this thesis. I should also thank the Health Research Board of Ireland for their generous support throughout my PhD. Similarly, I would like to thank those at the U.S. National Cancer Institute for working with the HRB to devising the Health Economics Fellowship programme that facilitated this research. Thanks go to Prof. Hugh Gravelle for his helpful comments on an early draft of Paper 1. Thanks also to Dr Ewout Steyerberg for his advice on the continuous time Markov models paper. A general thank you is owed to all at MGZ at Erasmus University for facilitating my enjoyable stay there for two years during my research, especially Prof. Dik Habbema and Dr Marjolein van Ballegooijen. Thanks are due to Prof. Charles Normand for his calm and attentive encouragement during the course of the thesis. Particular thanks are given to Dr Joost van Rosmalen, my thesis supervisor at Erasmus University. Joost gave many hours of dedicated assistance during the course of my thesis. I am deeply indebted to him.

CONTENTS

LIST OF ABBREVIATIONS

ACER	Average cost-effectiveness ratio
BIA	Budget impact analysis
CEA	Cost-effectiveness analysis
CEAC	Cost-effectiveness acceptability curve
CER	Cost-effectiveness ratio
CHB	Chronic hepatitis B
CIN	Cervical intraepithelial neoplasia
CTMM	Continuous-time Markov model
CUA	Cost-utility analysis
CVZ	College voor Zorgverzekeringen
DES	Discrete-event simulation
DNA	Deoxyribonucleic acid
DTMM	Discrete-time Markov model
ETV	Entecavir
EVPI	Expected value of perfect information
FIT	Faecal immunochemical test
HBV	Hepatitis B virus
HIQA	Health Information and Quality Authority
HPA	Health Protection Agency
HPV	Human Papillomavirus
HRB	Health Research Board
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IMF	International Monetary Fund
ISPOR	International Society for Pharmacoeconomics and Outcomes Research

LAM	Lamivudine
LYG	Life years gained
MGZ	Maatschappelijke Gezondheids Zorg
MISCAN	MIcrosimulation SCreening Analysis
NHB	Net health benefit
NICE	National Institute for Health and Clinical Excellence
PSA	Probabilistic sensitivity analysis
QALY	Quality adjusted life year
SMDM	Society for Medical Decision Making
SVR	Sustained virologic response
TDF	Tenofovir

INTRODUCTION

One of the key texts in the field of cost-effectiveness analysis (CEA) of healthcare interventions is Gold et al.'s 1996 *Cost-Effectiveness in Health and Medicine.* One of the opening chapters is dedicated to the question of how to frame a CEA [1]. In it Torrance et al. discuss how health economic analyses should be tailored to the policy choices they are to inform. Although the discussion is necessarily general as the range of issues is broad, it does provide a comprehensive overview of the importance of appropriate framing.

Despite the recognition of appropriate framing there remain examples of CEAs that are inappropriately specified for the policy question they are to answer. That is, the CEA as originally conceived does not best match the policy question under consideration. This thesis considers such examples, whereby CEAs may be reframed to better suit the particular policy question at hand.

Among the issues explicitly considered by Torrance et al. are the definition of the intervention itself, what comparator interventions to include, the intervention's target population and the time horizon of the analysis. The studies included in this thesis provide examples of how CEAs can be reframed to improve their fit to policy questions with regard to each of these issues.

This introduction describes the background to the use of models in CEA, including: a brief overview of the types of models used in CEA; an explanation as to why modelling is particularly useful in CEAs of cancer prevention; and, a brief consideration of why modelling may fail to influence policy in the way CEA practitioners might expect. The introduction then describes the aims of this thesis, before going on to provide an overview of each of the six studies contained in the thesis, explaining the background to each study and how it relates to the next. The discussion chapter at the end of the thesis summarises the contribution of each study and explains the relationship of each to the central theme of reframing models to correspond with policy questions.

Cancer and Secondary Prevention

Despite continuing advances in medicine, cancer still imposes a heavy burden on health globally. The International Agency for Research on Cancer estimated over 12.5 million incident cases and 7.5 million deaths worldwide in 2008 [2]. In Ireland, cancer accounted for over 8,000 deaths in 2007 [3]. Although treatment is improving, mortality from the disease remains high, with 5-year survival for all cancers combined being approximately only 50% in Europe [4]. In addition to the mortality burden, treatment itself can impose significant morbidity and survivors remain at risk of recurrence. Consequently, cancer still evokes fear among individuals and remains a priority for policy makers.

Efforts to control cancer include prevention. While primary prevention aims to reduce exposure to risk factors, secondary prevention through early detection of preclinical disease offers a means of reducing the incidence of late-stage, metastatic disease and improving cancer survival. Screening for cervical, breast and colorectal cancer have proved successful at reducing cancer morbidity and mortality in the 20th century and are now offered in many developed countries on a population basis. The 21st century promises further development of such screening programmes and advances in the detection of lung, oesophageal, prostate and other cancers.

The improvement of cancer screening will largely depend on trials of new tests and screening protocols and the identification of novel and more accurate risk markers. While trialling will remain essential to the appraisal of effectiveness of new interventions, modelling will also complement such efforts, especially as the design of optimal screening interventions has and will remain heavily dependent on modelling.

Considerable effort is devoted in the field of CEA to the improvement of modelling methods. This effort is to improve the reliability of CEA models as guides to healthcare policy in the allocation of scarce resources. This thesis aims to contribute to the improvement of modelling methods by investigating the use of models in CEAs of cancer prevention.

Models in Cost-Effectiveness Analysis

Buxton et al.'s definition of models as "a way of representing the complexity of the real world in a more simple and comprehensible form" captures the broad function of models [5]. The purpose of models in CEA is to inform decision makers in their healthcare resource allocation choices.

Models are used to simulate alternative intervention options and their anticipated consequences. Modelling includes: applying cost and quality-adjusted effects to clinical trial results to yield cost-effectiveness estimates; the extrapolation of results beyond trial follow-up periods; and, the combination and synthesis of data from multiple sources to compare interventions that have never been directly compared in clinical trials [6]. It is this last form of modelling that this thesis primarily concerns.

A number of authors have proposed taxonomies to describe the various features of models employed in CEA [7-10]. A simple type of model is the decision tree model, which is used to map out alternative treatment options and their expected consequences [11]. While suitable for simple decision problems, decision trees cannot account for time and quickly become unusable when repeated healthcare choices are considered [12]. State-transition models are better suited to time dependence and reoccurring choices and so have become widely used in CEA [13]. Such models allocate individuals or some portion of a cohort of patients to mutually exclusive health states and update state membership over time [8].

State-transition models can be analysed mathematically or using simulation methods [8]. Mathematical approaches have been used in the CEA literature [14-18]. While mathematical models have some advantages over simulation models, they can quickly become very complex as detail is added to the model, as a result they remain relatively little used in CEA [19]. The simulation approach to state-transition modelling offers greater ease of construction, more flexibility when altering model structure and tends to be more transparent; consequently, simulation-based state-transition models are now widely employed [13].

Simulation models may follow the progress of a homogenous cohort through a number of health states [7]. Alternatively, microsimulation or agent-based

modelling may be employed, whereby individuals pass through the model one at a time. Cohort simulation approaches have been popular within CEA, in part, as they can be easily run on desktop computers. They are commonly implemented as Markov models, whereby the Markov assumption assumes that transition probabilities out of the current state are independent of previous state membership [20]. In this sense, Markov models are described as "memoryless" [21]. This memoryless property is often not suitable for CEA modelling, as a patient's prognosis often depends on their prior health states. While it is possible to work around the Markov assumption by adjusting the structure of the model to account for contingent probabilities, this can result in state proliferation and the model can become large and cumbersome [22]. Such limitations of Markov modelling have furthered the development of microsimulation methods.

While the term "cohort model" is often used to describe an aggregate-level rather than individual-based simulation, it can also refer to models that simulate one group of individuals, as opposed to multi-cohort models that follow more than one cohort [9, 23]. The single cohort in a cohort model may be of a common birth year, as is often the case in models of screening, as the start of the intervention is not determined by disease onset, but rather the age attained by the individuals. Conversely, in examples where the intervention does begin with the onset of disease, the cohort may simply be a cohort of patients diagnosed in a common year and so may be made up of individuals of mixed ages. Consequently, it can be misleading to imagine the individuals within a cohort model as homogenous.

Another distinction is between static and dynamic models [10]. Static models do not allow for interaction between agents while dynamic models do. Dynamic models can be used to simulate the spread of infectious diseases [24]. Dynamic models tend to be more complex and require more data than static models. Typically cohort models are static and dynamic models are multi-cohort models. However, pseudo-dynamic models can be used to simulate infection dynamics in static models [25].

A further distinction between model types relates to the treatment of time. In discrete time models state membership changes at discrete intervals, such as months or years. While the discrete-time approach is the most commonly applied in CEA

[26], it can only approximate the timing of real events, since they do not occur at discrete intervals, but in continuous time [20]. Consequently, continuous time models have been applied in CEA, including discrete event simulation, which also relaxes the assumption of mutually exclusive health states [22, 27].

One distinction less frequently mentioned in reviews of model methodology is that between what are described variously as shallow or deep, epidemiological or biological, or empirical or theoretical models [28-30]. Shallow models are those that reproduce empirically observable disease states without modelling the underlying biological processes of disease and intervention, while deep models do attempt to model such processes [29]. Shallow models may be easier to construct and require less data, but they can lack the flexibility to illustrate a wide range of policy alternatives [28].

The principal model used in this thesis is the MIcrosimulation SCreening ANalysis (MISCAN) model developed at Erasmus Medical Centre at Erasmus University Rotterdam, the Netherlands [31]. MISCAN is a static, continuous time, discrete event, microsimulation model that simulates the underlying disease progression of cancer and the effects of early detection and treatment. MISCAN can simulate single or multiple birth cohorts over finite or open-ended time periods. MISCAN has been applied to, breast, cervical, colorectal, lung and prostate cancer screening and vaccination against HPV.

CEA and the Cost-Effectiveness Plane

Estimates of costs and effects from CEA models are typically represented in the cost-effectiveness plane. The cost-effectiveness plane is a two dimensional graph that plots net costs of the intervention on the vertical axis against the net health effects, often measured in quality adjusted life years gained (QALYs) or life years gained (LYG).

The origin of the cost-effectiveness plane often represents either the current standard of care or the do-nothing option where no intervention is offered. Cost-effectiveness estimates that lie in the north-west quadrant of the graph are strictly not preferred, as they involve more costs and worse health outcomes. Estimates in

the south-east quadrant are strictly preferred, as these involve reduced costs and improved health outcomes. Most commonly, cost-effectiveness estimates will lie in the north-east quadrant, where costs are increased, but health outcomes are also improved. Estimates in the south-west quadrant represent reduced costs and worse health outcomes.

When multiple alternatives are compared we must consider the possibility of dominance. Simple or strong dominance occurs when an intervention lies to the north-west of another intervention; being more costly and less effective. Extended or weak dominance occurs when an intervention lies to the north-west of a line connecting two other options in the cost-effectiveness plane; being more costly and less effective than some average of the two other interventions.

The set of strategies that are not subject to strong or weak dominance make up the efficient set of interventions. The line joining these strategies is the cost-effectiveness frontier. The average cost-effectiveness ratio (ACER) is calculated by dividing the incremental costs of each intervention relative to the origin by the incremental effects. More important than the ACER is the incremental cost-effectiveness ratio (ICER). The ICER is the ratio of the incremental cost and effects of an efficient intervention relative to the preceding less costly efficient intervention.

The ICER is the principal measure of cost-effectiveness used in CEA. Interventions' ICERs are compared to the cost-effectiveness threshold, which represents the maximum willingness to pay for a unit of health gain. Those interventions with ICERs below the threshold are considered cost-effective, although it should be noted that negative ICERs have no useful interpretation and should be disregarded. Among a mutually exclusive set of interventions for the same condition, the intervention with the highest ICER under the threshold is considered optimal, in that it will yield the largest health gain at an acceptable level of cost-effectiveness for all interventions considered.

Features of Cancer Prevention Relevant to CEA Modelling

The studies in this thesis all concern cancer prevention, primarily using secondary prevention through screening for preclinical disease. Prevention in general and screening in particular have a number of noteworthy features that make modelling particularly relevant for CEA. These features are noted here as they provide relevant context to the studies contained in the thesis.

The intensity of screening can be varied in many cases. This can be achieved by varying screening intervals, start and stop ages, numbers of follow-up tests in response to equivocal primary test results and quantitative metrics in some tests. There is typically some sort of dose-response relationship in screening interventions, with more frequent screening being associated with more protective effect of detecting earlier disease. Similarly, more intense screening will usually result in greater net costs.

This variation of costs and health effects with screening intensity creates scope for considerable variation in screening cost-effectiveness. While screening can be highly cost-effective at low intensities, the same screening method may be highly inefficient at high intensities due to diminishing marginal effectiveness of increased screening. For example, Eddy estimates an ACER of cervical screening between ages 20 and 47 every four years to be approximately $10,000/ LYG, while annual screening over the same age range has an estimated ICER in excess of $1,000,000/LYG [32]. Consequently, policy makers need to ask not only is screening cost-effective, but also at what intensity.

This variation of cost-effectiveness with screening intensity means trials typically cannot identify the optimal intensity, as it is infeasible to compare all possible screening intensities in a trial [28]. Even if a trial indicates a given screening strategy is beneficial, it cannot provide evidence of the incremental benefits relative to other screening strategies, unless those alternative strategies have also been trialled [33]. Consequently, modelling is necessary to combine trial and other available data together to simulate the wide range of screening alternatives to determine the efficient frontier of screening strategies and the ICERs between them [34].

Modelling is also required in the case of screening because trials could take decades to report final outcomes. The requirement for preventative services in the interim means it is infeasible to wait many years to determine the optimal policy. Furthermore, if screening is already provided this may preclude trialling, as individuals are unlikely to accept trial protocols involving screening at lower intensities than at current provision on the presumption this would offer less protection.

Brown and Buxton note that since screening is given to asymptomatic individuals it is free from the 'rule of rescue' [6]. This suggests that screening is not subject to the tendency to mandate greater health spending on those already sick, irrespective of how much health gain is achieved as a result of resources being allocated to them. Consequently, Brown and Buxton suggest this makes screening particularly suitable for CEA, since decisions regarding the intervention can be made dispassionately on the basis of expected costs and health effects. However, the ongoing debate over the effectiveness and cost-effectiveness of breast and prostate cancer screening indicate that screening policy is not always free from controversy or strong feelings [35, 36].

A particular feature of screening is that it does not improve health itself, but as a secondary preventative measure it potentially enhances treatment outcomes by facilitating treatment at an earlier stage of disease through early detection [33]. Consequently, the cost-effectiveness of screening is contingent on the cost-effectiveness of treatment at different stages of disease progression.

Not only does screening not enhance health itself, but it can lead to harm, directly or indirectly [37]. An example of a direct harm is perforation during colonoscopy [28]. Indirect harms include increased cancer risk following repeated radiation exposure during mammographic screening, unnecessary biopsies and treatment following false positives, as well as unnecessary treatment resulting from overdiagnosis of indolent disease [28]. Furthermore, screening can lead to a false sense of protection or reassurance, as screenees may be unaware of the possibility of false negative screen results [38]. The potential harms of screening per screenee may be small relative to the potential gains for each screen detected cases, however the aggregate balance of benefits and harms over all screenees have to be considered in CEA.

As screening is provided to asymptomatic individuals, the total number of screened individuals in a screened population can be very large, while the number of screen detectable preclinical cases may be very small. Consequently, the cost-effectiveness of screening can be highly sensitive to disease prevalence and the expected health gain of early treatment. Similarly, the cost-effectiveness of screening is typically quite sensitive to the cost of screening and to differences in the relative performance of alternative screening technologies.

The cost-effectiveness of preventative interventions is typically particularly sensitive to intertemporal discounting [39]. Since prevention typically involves expenditure in the present that will not yield health gains and cost savings until the future, discounting often leads to far less favourable cost-effectiveness ratios than undiscounted outcomes [40].

The cost-effectiveness of screening can vary between subgroups. The cost-effectiveness of prevention varies with disease incidence, being more cost-effective with higher incidence, ceteris paribus or all else equal. Similarly, the cost-effectiveness of prevention will diminish with reduced remaining life expectancy. As the incidence of many diseases increases with age but remaining life expectancy diminishes, there are two countervailing effects on the cost-effectiveness of prevention with age. Disease incidence can also vary with factors such as health behaviours or genetic predisposition. Consequently, part of the concern with the cost-effectiveness of screening is to determine which subgroups should deviate from the screening intensity offered to the average risk population [16].

In principle, to achieve optimal cost-effectiveness, interventions that exhibit increasing ICERs with increasing intervention intensity should be increased in intensity until the ICER reaches the cost-effectiveness threshold. However, population screening can have a very large overall budget impact or other resource requirements. Therefore, decision makers may choose a screening intensity with a low ICER relative to the threshold in order to moderate the programme's budget impact.

A good example of high budget impact interventions being chosen at intensities with low ICERs is colorectal screening. Ireland and the Netherlands recently opted to provide colorectal screening strategies with ICERs estimated at €3,200/QALY and €3,900/QALY respectively, while the commonly quoted cost-effectiveness thresholds in Ireland and the Netherlands at the time were €45,000/QALY and €20,000/QALY respectively [41-44]. Consequently, these examples show that while it is common in CEA to assume the budget impact of an intervention is small and that reimbursement decisions can solely be made with regard to cost-effectiveness ratios [45], the budget impact of screening can be sufficiently large to cause a departure from the conventional intervention optimisation framework.

The Limits to the Impact of Modelling on Policy

A number of commentators have noted that CEA modelling appears to have made only a modest impact on policy [46-51]. This has prompted others to consider why this might be the case.

At a fundamental level, a weak societal acceptance of CEA may explain its modest impact. It has been questioned if CEA is consistent with society's preferences for the allocation of health [52]; more specifically, others have questioned some of the methods central to CEA, including the use of QALYs, aggregation of health outcomes over individuals and discounting [53-57]; it has also been suggested that dissatisfaction with CEA is a manifestation of a broader dissatisfaction with healthcare rationing implicit within CEA[51, 58]. Certainly a weak societal acceptance of CEA undoubtedly explains part of its modest policy impact, but there are also other issues for consideration.

The legal and institutional context of CEA can also partly explain its lack of impact in certain countries. For example, the impact of CEA on healthcare resource allocation may be expected to be modest in the US where there are legal restrictions on the use of cost-effectiveness metrics to determine eligibility for care [59]. Similarly, CEA's impact is likely to be limited in Germany, where it is illegal to withhold care on the basis of cost [60]. Conversely, CEA may be expected to be relatively more influential in England and Wales, as there is a legal requirement that

interventions recommended for use by the National Institute for Health and Clinical Excellence (NICE) be provided to patients [61].

It has been suggested that CEAs may have a limited impact on policy because decision makers do not trust their findings [62]. Such lack of faith in CEA has been attributed to decision makers perceiving CEA as representing poor quality evidence [49]. Part of this perception may be due to the observed failure of many analyses to adhere to methods guidelines [63, 64]. Mistrust in CEA can also stem from a lack of transparency of modelling, leaving the impression of the 'black box' [65]. Concerns of lack of transparency are heightened by the potential for bias, especially as financial interests are often at stake [66, 67].

Even if decision makers trust the findings of CEA models there may be other reasons for CEA to have a modest policy impact. CEA may not fit readily with decision maker priorities. It has been suggested that decision makers are typically concerned with short-term budget impact from the payer's perspective rather than the long-term outcomes from a societal perspective [68]. It is also noted that decision makers often face inflexible budget allocations and cannot direct resources to cost-effective interventions, so CEA is of little use [49, 51, 68].

Finally, the weak influence of CEA has also been attributed to a lack of comprehension of modelling by decision makers [51, 52]. Bryan et al. and Williams et al. both describe how decision makers felt an inability to reliably interpret the results of cost-effectiveness models [46, 69]. Similarly, Hoffman et al. report that decision makers often appear to have limited understanding of the underlying concepts and methodology of CEA and require support in interpreting evidence [49].

In sum, previous research has identified an apparently lack of impact of CEA on policy choices and has proposed a number plausible explanations for why this might be so.

Aims of this Thesis

This thesis aims to describe and investigate cases in which CEAs could be reframed to better correspond to the policy questions they are to inform. It intends to illustrate how relatively simple but not always obvious aspects of model structure can result in poor fit to the policy choices and, similarly, how simple steps can be taken to adjust or reframe the CEA to better inform decision makers. The purpose of this investigation is to prompt some critical reflection on the part of decision makers and analysts regarding the appropriateness of the models and methods they use given the policy choices in question. By drawing attention to the importance of model fit to policy choices and illustrating the scope for beneficial reframing, this thesis hopes to contribute to the ongoing process of methodological development within the field of CEA.

INTRODUCTION TO PAPERS 1 - 6

This introduction now turns to the studies included in the thesis. The background to each paper is outlined here. In each case, the gap in the existing literature that each study is intended to address is described. In addition, an explanation of the link between each successive study is also provided.

Background to Papers 1 & 2

The first two papers presented in this thesis concern differential discounting of costs and health effects in CEA. Consequently, the background to both papers is presented together. The first paper was published in Value in Health in 2011 [70]. It identified a methodological problem when applying differential discounting. The second paper responds directly by proposing a solution to the problem identified in the first and has been submitted for review.

Discounting in CEA

Discounting is employed in economic analyses to account for positive time preference. Positive time preference is the observed tendency of individuals to prefer receiving economic goods sooner rather than later. Time preference is given three principal theoretical justifications [71]. The first is that future consumption is less certain than current consumption, therefore current consumption is preferred. The second is that expectations of income growth coupled with diminishing marginal utility of consumption provides a rationale for shifting consumption forward in time. The final justification is simple human impatience for enjoying good things sooner.

The rationale for including discounting in economic analyses is that such analyses should reflect the preferences of society, including time preferences [72]. Further rationale is provided by interest on borrowing and loans. As positive interest rates imply a real cost of borrowing and an opportunity cost of spending rather than saving, economic analyses also need to include these costs [73].

Economic analyses typically employ the exponential or constant discounting model, whereby values are discounted at a constant rate over time, generally at the same rate for all goods. The constant discounting model has the theoretical and practical

advantages of being derived from a model of rational consumer behaviour and being convenient to apply [74].

Despite the widespread use of exponential discounting, its deficiencies as a representative model of human behaviour have been noted [74, 75]. While a model of rational consumer behaviour suggests individuals should have equal time preferences for all tradable goods, as any differences in time preferences would be arbitraged away [75], evidence suggests individuals do have different time preferences for different goods [76]. Furthermore, empirical evidence has shown that actual time preferences do not conform to the exponential model and can be better represented with alternative functional forms that feature discount rates that decline over time [77, 78].

Despite the failure of exponential discounting to represent empirically observed preferences, the model remains the consensus means for accounting for time preferences in economic analyses [74]. Regarding the appropriate discount rate, common practice is to use the rate of return on government bonds as a measure of the riskless rate of return as an indicator of the appropriate societal discount rate for publicly funded projects [79].

The application of discounting in the economic appraisal of healthcare is controversial as it tends to result in less favourable cost-effectiveness ratios than otherwise, especially in preventative care [80]. Somewhat understandably this has led to a number of critiques of discounting, often expressed with a notable sense of annoyance or dissatisfaction with current practice [54, 81, 82].

The literature features a number of arguments for zero discounting, non-constant discounting and other alternative forms of discounting that provide more favourable cost-effectiveness estimates [54, 56, 77, 83-85]. While much of these calls for alternative discounting models convey an impression of special pleading and have made little impact on discounting guidelines, the case of differential discounting has been more successful.

Differential Discounting

Healthcare is considered a superior good in economic terminology, meaning that as societies get richer, they typically spend an increasing proportion of income on healthcare [86]. The reason being that the marginal utility of health is anticipated to decrease by less than that of consumption as income rises and basic material needs are met [87]. The consequence of the marginal utility of health falling by less than that of consumption with income growth is an increase in the consumption value of health; i.e. an increasing willingness to pay for healthcare [88].

A number of health economists have argued that an appropriate way to account for such a growing value of health in CEA is to apply a lower discount rate to health effects than to costs [88-94]. However, not all health economists agree and the claim for differential discounting has been debated vigorously [73, 95, 96].

Differential discounting has been adopted in a number of countries to date. NICE employed differential discounting with rates of 6 & 1.5% for costs and health effects respectively until its 2004 methods guidelines revised the discount rate to 3.5% for both costs and effects [93]. More recently, NICE amended its guidelines to require differential discounting at 3.5 & 1.5 % in the case of interventions that are substantial in restoring health for sustained periods of 30 years or more [97]. The Netherlands, Belgium and Poland all require differential discounting at 4 & 1.5%, 3 & 1.5% and 5 & 3.5% respectively [98-100].

Despite the long-running theoretical debate over differential discounting, a degree of consensus was achieved among some of those involved in a recent paper by Claxton et al. [101]. It shows that the appropriate form of discounting depends on the decision rules used in CEA and whether health or welfare is the maximand. In particular, it shows that if the healthcare budget is fixed and a cost-effectiveness threshold is used to determine whether an intervention is cost-effective, then the differential between the discount rate on costs and health effects should approximate the annual growth rate of the threshold. They also show that under such circumstances, an increase in the value of health requires an equal reduction in both the discount rate for costs and health effects.

Claxton et al.'s contribution to the debate over differential discounting is profound in that it shows that a change in the threshold, not a change in the value of health is what determines if differential discounting should be used. This is a useful result, as the cost-effectiveness threshold is arguably a far more tangible concept than the value of health. Indeed, cost-effectiveness thresholds have been explicitly stated in the case of NICE's threshold range [102]. Alternatively, an implied threshold can, in principle, be observed from the resource allocation choices in a budget constrained health system [103]. While most thresholds are not explicit, there has been little evidence of nominal or real threshold growth [104]. The UK threshold range has remained unchanged over a number of years [102], while the notional threshold used in Ireland was recently more than halved [105].

Although most of the countries employing differential discounting adopted the practice before the publication of Claxton et al., it is useful to consider the implied rates of threshold growth. For example, the former NICE differential rates imply an approximate annual rate of real threshold growth of 4.5%, which certainly seems unsustainable. Indeed, the approximate rate of growth of 2.5% implied by the Dutch rates might also prove unsustainable. Consequently, the insight into appropriate differentials offered by Claxton et al. should prompt consideration if differentials need to be reduced to correspond with more plausible rates of threshold growth.

In summary, differential discounting has been subject to robust theoretical debate to date and there still remain some theoretical issues unresolved regarding differential discounting [106, 107]. Despite this, it has become recommended practice in a number of countries.

Paper 1: The Problem of Comparability under Differential Discounting

The existing literature on differential discounting has primarily addressed whether the application of differential discounting is appropriate and what is its theoretical justification. Little consideration has been given to the practical application of differential discounting and its implications for decision making when applied to the variety of policy questions posed in CEA. In particular, little consideration has previously been given to the consequences of differential discounting for models employing multiple cohorts that start an intervention over a range of time periods.

The most notable existing consideration of the effect of differential discounting on cohorts starting interventions at different time periods is the well-recognised postponing paradox. The postponing paradox was first described by Keeler and Cretin [95]. It is based on the observation that the cost-effectiveness of any intervention will improve if its implementation is postponed from the current cohort of patients and given to the following year's cohort instead. According to the paradox, decision makers seeking the best cost-effectiveness ratios have an incentive to continuously postpone implementation, thus never actually choosing to provide the intervention.

The paradox has been convincingly dismissed as irrelevant to actual policy choices on the grounds that decision makers use CEA to determine the choice of optimal intervention at any given point in time, rather than determining the optimal timing of interventions [89]. However, previous authors have not considered how the postponing paradox might apply to comparisons between models with different numbers of future cohorts.

A more recent study has recognised the implications of differential discounting for the cost-effectiveness estimates when multiple future cohorts are added. Hoyle and Anderson recognised that the inclusion of future cohorts would drive an intervention's ICER down, all else equal [108]. However, that study did not consider how differential discounting could compromise comparisons between analyses, but rather considered the effect of reducing ICERs a rationale for including all an intervention's future recipient cohorts into a CEA.

In summary, while the existing literature recognises that differential discounting will reduce an intervention's ICER the later the recipient cohort receives the intervention relative to the discount year, it has not yet been considered whether this presents any problem for comparing interventions. The first paper in the thesis addresses this gap, using the example of a multi-cohort model of HPV vaccination in the Netherlands. It shows how the cost-effectiveness of vaccination increases when more cohorts are added to the model.

Paper 2: Solving the Problem of Comparability

The second paper in the thesis is intended to present a solution to the problem described in the first paper. Since the problems of comparing cost-effectiveness estimates from models with differing numbers of future cohorts was only recently described in the literature by the first paper in the thesis, there is no clear gap in the literature regarding potential solutions to the problem as there previously was no recognised need for a solution. Rather, the second paper sits in the context of a complete absence of any literature at all addressing possible solutions to the problem of comparability.

Despite the lack of previous literature regarding solutions to the problem of comparability, it is worth noting one relevant aspect of the existing literature. A simple solution to the problem of comparability described in the first paper would be to apply a two-step discounting procedure similar to that previously suggested to reconcile differences between private and social discount rates [79, 109]. In the case of differential discounting, the first discounting step would be to discount the cost and effects differentially for each cohort to the year it starts the intervention. The second step would be to discount all costs and effects from each intervention start year back to a common discount year at an equal discount rate.

A minor problem with the two-step solution is that it is not obvious what rate the equal discount rate for the second step should be. A more fundamental problem occurs in interventions with shared effects. The presence of shared health benefits between cohorts as a consequence of herd immunity means there will be cost

savings and health effects that cannot be attributed to any one cohort when multiple recipient cohorts receive the intervention. Therefore, in the first discounting step there will be no separate identifiable intervention start years to discount to. Consequently, two-step discounting cannot be applied in the case of infectious diseases. The challenge therefore is to find a general solution to the problem of comparability that does not require the identification of specific discount years for shared benefits.

Paper 3: The Implementation Time Horizon

Background

Paper 3 considers what is described as the implementation time horizon. Time horizons are the point until which costs and effects are assessed in CEA models. Time horizons are well understood within CEA and it is recognised that they should be sufficiently long to capture all relevant costs and effects of an intervention [1], which often requires the assessment of costs and effects until death [13]. However, there is another practice sometimes used in models whereby, although the assessment of costs and effects continues until death, the implementation of the intervention is assumed to cease at a certain point in time in the model. This is what is described as the implementation horizon.

No previous study has described the implementation time horizon or assessed its effects. As such, there is a clear gap in the literature regarding this aspect of model structure.

Many models used in CEA only simulate one cohort of recipients [110-112]. In the case of screening interventions this cohort is typically of a common age. While modelling one cohort of intervention recipients alone has the benefits of simplicity and transparency, there are often good reasons to include multiple cohorts of different birth years [110, 112]. Some multi-cohort models feature implementation time horizons, including models of screening programmes.

The use of implementation time horizons in multi-cohort models prompts the concern that they could yield results that are unrepresentative of actual

implementation. The basic concern was that it is unrealistic to assume that all screening would cease for all cohorts after a finite time span and that the results from such a model could be unrepresentative of reality, as screening is likely to be maintained. The basis for this concern was that not all screens may be equally cost-effective, therefore the screening moments censored by the implementation horizon may have different cost-effectiveness to those within the horizon. Consequently, the screening strategies identified as optimal by the model with an implementation horizon could be different from those by a model in which cohorts do not finish screening prematurely.

The absence of previous literature appraising the influence of time horizons and the concern of non-optimal choices from such models motivated the investigation described in Paper 3. It uses a previously published multi-cohort model of colorectal screening used to guide screening policy in the Netherlands that employed a 30-year implementation horizon. The analysis investigates the consequences of relaxing the 30-year time horizon, as well as the effects of shorter horizons of 10 and 20 years.

Paper 4: Aggregation in Multi-Cohort Models

The model considered in Paper 3 is an example of a multi-cohort model of screening. While, Paper 3 questioned whether a finite implementation horizon was appropriate, it did not consider if the aggregation of costs and health effects over all cohorts together is most appropriate. Paper 4 examines the aggregation of cost and effects in multi-cohort models.

The introduction of or change to a long-duration screening programme will affect multiple cohorts: the cohort at the screening start age, all the older cohorts within the screening age range and future incident cohorts. There are few papers in the literature that explicitly address the appropriate model structure regarding the numbers of cohorts to include. One such study, Dewilde and Anderson notes that the majority of CEAs of screening interventions only model the current incident cohort [110]. It describes how such models fail to include the prevalent cohort, and consequently may fail to accurately reflect the cost-effectiveness of actual implementation. Dewilde and Anderson uses the example of the cost-effectiveness

of cervical screening, showing that intensifying screening is less cost-effective when their model is extended to include prevalent cohorts.

A related study by Hoyle and Anderson recognises that if prevalent and incident cohorts differ in their cost-effectiveness, then the inclusion of future incident cohorts will also affect the cost-effectiveness estimates of multi-cohort models [108]. Consequently, it advocates including all intervention recipient cohorts, both present and future.

These studies noting the potential differences in cost-effectiveness between prevalent and incident cohorts are correct in their recognition of heterogeneity between cohorts. However, they give little consideration as to whether it is appropriate to aggregate cohorts together in CEA models, or whether cohorts with different cost-effectiveness estimates should be treated as separate subgroups. This is despite the fact that the principle of considering subgroups of different cost-effectiveness separately has long been recognised in CEA [1, 113].

In summary, the literature on aggregation of cost-effectiveness estimates across cohorts, especially in the case of CEAs of screening programmes, has given inadequate consideration to the potential problems that might arise. Indeed, it is notable that the most relevant existing literature on the topic explicitly advocates methods that run counter to the principle of subgroup analysis and that do not facilitate optimal decision making when separate decisions can be made for cohorts individually. As a result, there was a notable gap in the literature for a study articulating the disadvantages of aggregating across cohorts. Paper 4 has since been published in the Journal of Medical Decision Making [114].

Paper 5: Adjusting the Optimal Screening Interval for Disease Incidence

Background

The previous paper, Paper 4 considered the issue of subgroup disaggregation in order to identify an optimal screening strategy. Paper 5 also considers the issue of subgroup differentiation in order to identify an optimal screening strategy. It considers the optimal intensification of breast cancer screening for subgroups with incidence higher than the average risk population.

The primary focus of Paper 5 is the demonstration of a mathematical model to estimate the change in the optimal screening interval with disease incidence relative to the optimal screening interval for the average risk population.

While most screening models currently use simulation methods, analytical mathematical methods models have also been used. One of the earlier mathematical models in the literature is Kirch and Klein [14]. Kirch and Klein use a simple model of screening to show that the optimal screening interval varies proportionately to square root of the incidence rate. While the model used is of limited realism owing to its assumptions of perfect screen sensitivity and constant duration of the preclinical screen detectable disease state, the concept of relating the optimal screening interval to the disease incidence is useful.

While a number of other more detailed mathematical models of optimal screening have been developed subsequent to Kirch and Klein including work by Zelen and also Parmigiani [15-18, 115, 116], none have so explicitly considered what the relationship between disease incidence and the optimal screening interval might be. Furthermore, the mathematical complexity that arises in such models when additional realism is added has led to the development of models that are not analytically solved, but rather employ simulation. The move to simulation methods means that explicit relationships between parameters and policy variables such as that identified by Kirch and Klein in their simplified model are less likely to be considered. As such, the literature lacked an analysis that considered the incidence-optimal interval relationship in a model with sufficient realism to be credible.

Paper 5 is based on a manuscript previously submitted to the Journal of Medical Decision Making and rejected. That manuscript had been initially drafted by N Mushkudiani and JDF Habbema. It was further developed by F Goudsmit and EW Steyerberg and submitted. The rejected manuscript and associated model were then substantially revised by J O'Mahony and J van Rosmalen and has been resubmitted to the Journal of Medical Decision Making and is under review.

The major revisions include the addition of a verification of the mathematical model against a simplified version of the MISCAN breast screening model programmed to match the mathematical model. The validation revealed errors in the mathematical model, which required revision. A validation of the mathematical model was conducted against a complete version of the MISCAN breast screening model used in a recently published CEA of breast screening in Switzerland [117]. The model of breast screening used for the validation exercise was an extended version of the Swiss CEA, as that analysis only compared ad-hoc screening to one organised screening strategy for average-risk women, whereas the validation required the comparison of a range of screening intervals over a range of incidence rates. Other changes to include the addition of imperfect test specificity to the model and the addition of a literature review.

The coding of the simplified MISCAN model revisions to the mathematical model were undertaken largely by J van Rosmalen, while the analysis with the microsimulation models including the comparison to the mathematical model and literature review were largely undertaken by J O'Mahony.

Paper 6: The Relevance of Alternative Screening Strategies to CEAs of HPV Vaccination

The previous paper, Paper 5 features an analysis of the cost-effectiveness of screening in which estimates of cost-effectiveness are made with reference to the cost-effectiveness frontier. The frontier in Paper 5 is estimated with a mathematical model as a continuous function of the screening interval. Paper 6 considers the relevance of the cost-effectiveness of alternative cervical screening strategies for cost-effectiveness estimates of HPV vaccination. Like Paper 5, it considers the cost-effectiveness frontier of alternative screening options. In this case, the frontier joins

discrete alternatives of cervical screening estimated from a microsimulation model. The purpose of considering many screening alternatives in Paper 6 is to ensure the frontier is estimated as completely as reasonably possible.

Variation in screening intensity will lead to variation in cost-effectiveness. An early study estimating changes in ICERs of cervical screening with the screening intervals is Eddy [32]. A more recent study considering alternative screening start and stop ages as well as different intervals is van den Akker-van Marle et al. [118]. Both these studies present cost-effectiveness frontiers for a wide variety of screening intensities and estimate low ICERs for low intensity screening and very high ICERs for high intensity screening. These studies demonstrate the relevance of considering alternative screening intensities when appraising screening cost-effectiveness.

The successful development of two vaccines against HPV by two manufacturers marked a significant development in the prevention of cervical cancer. The advent of HPV vaccination prompted a large number of CEAs assessing the vaccines' cost-effectiveness. Following appraisal, HPV vaccination has been adopted in many developed countries to date.

An important part of the appraisal of HPV vaccination relates to changes in screening. Most analyses assume that the current screening strategy will be maintained and estimate the ICER of adding vaccination to that strategy [119]. However, a number of analyses recognise that screening may need to change for vaccinated women, as a reduction in disease incidence will erode the cost-effectiveness of the current screening strategies and some reduction in intensity may be necessary. This is likely to be especially true in settings where recommended screening is very frequent.

The failure of many analyses to include alternative screening strategies is despite a well-recognised principle in CEA that all reasonable comparators should be considered [1]. It is known that the inclusion of relevant alternatives means that comparators should not be restricted to current practice, as current practice may

itself not be efficient [64]. Omitting other cost-effective comparators can lead to a more favourable cost-effectiveness estimate of the intervention of interest.

There is no paper within the cost-effectiveness literature that has considered the implications of considering all screening alternatives when assessing the cost-effectiveness of vaccination as opposed to simply adding vaccination to the current screening strategy, either from a theoretical perspective or with an empirical analysis that simulates the various alternative strategies. Consequently, there was a notable gap in the literature for a study that documents the comparisons made in the literature to date and investigates the implications for considering alternative screening strategies.

REFERENCES

1. Torrance GW, Siegel JE, Luce BR. Framing and Designing the Cost-Effectiveness Analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 54-81.

2. Thun MJ, DeLancey JO, Center MM, Jemal A, Ward EM. The global burden of cancer: priorities for prevention. Carcinogenesis. 2010;31(1):100-10.

3. NCRI. Cancer in Ireland 2011. Cork: National Cancer Registry of Ireland, 2011.

4. Verdecchia A, Francisci S, Brenner H, Gatta G, Micheli A, Mangone L, et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. Lancet Oncology. 2007;8(9):784-96.

5. Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life. Health Economics. 1997;6(3):217-27.

6. Brown J, Buxton M. The economic perspective. British Medical Bulletin. 1998;54(4):993-1009.

7. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Economics. 2006;15(12):1295-310.

8. Mandelblatt JS, Fryback DG, Weinstein MC, Russell LB, Gold MR, Hadorn DC. Assessing the Effectiveness of Health Interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 135-75.

9. Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. The Journal of the Operational Research Society. 2007;58(2):168-76.

10. Kim S-Y, Goldie SJ. Cost-effectiveness analyses of vaccination programmes: A focused review of modelling approaches. PharmacoEconomics. 2008;26(3):191-215.

11. Sonnenberg FA, Roberts MS, Tsevat J, Wong JB, Barry M, Kent DL. Toward a Peer Review Process for Medical Decision Analysis Models. Medical Care. 1994;32(7).

12. Beck JR, Pauker SG. The Markov process in medical prognosis. Medical Decision Making. 1983;3(4):419-58.

13. Sculpher M, Fenwick E, Claxton K. Assessing Quality in Decision Analytic Cost-Effectiveness Models: A Suggested Framework and Example of Application. PharmacoEconomics. 2000;17(5):461-77.

14. Kirch RLA, Klein M. Examination schedules for breast cancer. Cancer. 1974;33(5):1444-50.

15. Parmigiani G. On Optimal Screening Ages. Journal of the American Statistical Association. 1993;88(422):622-8.

16. Zelen M. Optimal scheduling of examinations for the early detection of disease. Biometrika. 1993;80(2):279-93.

17. Baker R. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Management Science. 1998;1(2):103-13.

18. Lee SJ, Zelen M. Scheduling Periodic Examinations for the Early Detection of Disease: Applications to Breast Cancer. Journal of the American Statistical Association. 1998;93(444):1271-81.

19. Karnon J, Goyder E, Tappenden P, McPhie S, Towers I, Brazier J, et al. A Review and Critique of Modelling in Prioritising and Designing Screening Programmes. Health Technology Assessment. 2007;11(52):1-166.

20. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Medical Decision Making. 1993;13(4):322-38.

21. Briggs A, Sculpher M. An Introduction to Markov Modelling for Economic Evaluation. PharmacoEconomics. 1998;13(4):397-409.

22. Caro JJ. Pharmacoeconomic Analyses Using Discrete Event Simulation. PharmacoEconomics. 2005;23(4):323-32.

23. Habbema JDF, Boer R, Barendregt JJ. Chronic Disease Modeling. In: Heggenhougen K, editor. International Encyclopedia of Public Health. Oxford: Academic Press; 2008. p. 704-9.

24. Brisson M, Edmunds WJ. Economic Evaluation of Vaccination Programs: The Impact of Herd-Immunity. Medical Decision Making. 2003;23(1):76-82.

25. Bauch CT, Anonychuk AM, Van Effelterre T, Pham BZ, Merid MF. Incorporating Herd Immunity Effects into Cohort Models of Vaccine Cost-Effectiveness. Medical Decision Making. 2009;29(5):557-69.

26. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Economics. 2002;11(1):33-42.

27. Caro JJ, Möller J, Getsios D. Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations? Value in Health. 2010;13(8):1056-60.

28. Knudsen AB, McMahon PM, Gazelle GS. Use of Modeling to Evaluate the Cost-Effectiveness of Cancer Screening Programs. Journal of Clinical Oncology. 2007;25(2):203-8.

29. Feuer EJ, Etzioni R, Cronin KA, Mariotto A. The use of modeling to understand the impact of screening on US mortality: examples from mammography and PSA testing. Statistical Methods in Medical Research. 2004;13(6):421-42.

30. Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? The European Journal of Health Economics. 2003;4(3):143-50.

31. Habbema JDF, van Oortmarssen GJ, Lubbe JTN, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine. 1985;20(1):79-93.

32. Eddy DM. Screening for Cervical Cancer. Annals of Internal Medicine. 1990;113(3):214-26.

33. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics. 1984;40(1):1-14.

34. Pignone M, Ransohoff DF. Cross-Model Comparisons to Improve the Value of Modeling. Medical Decision Making. 2011;31(4):524-6.

35. Barry MJ. Screening for Prostate Cancer: The Controversy That Refuses to Die. New England Journal of Medicine. 2009;360(13):1351-4.

36. Kopans DB. The Breast Cancer Screening Controversy and the National Institutes of Health Consensus Development Conference on Breast Cancer Screening for Women Ages 40-49. Radiology. 1999;210(1):4-9.

37. Russell LB. Exploring the Unknown and the Unknowable with Simulation Models. Medical Decision Making. 2011;31(4):521-3.

38. Marshall KG. Prevention. How much harm? How much benefit? 1. Influence of reporting methods on perception of benefits. Canadian Medical Association Journal. 1996;154(10):1493-9.

39. Luce BR, Simpson K. Methods of cost-effectiveness analysis: areas of consensus and debate. Clinical Therapeutics. 1995;17(1):109-25.

40. Westra TA, Parouty M, Brouwer WB, Beutels PH, Rogoza RM, Rozenbaum MH, et al. On Discounting of Health Gains from Human Papillomavirus Vaccination: Effects of Different Approaches. Value Health. 2012;15(3):562-7.

41. HIQA. Health technology assessment of a population-based colorectal cancer screening programme in Ireland. Dublin: Health Information and Quality Authority, 2009.

42. Wilschut JA, Hol L, Dekker E, Jansen JB, van Leerdam ME, Lansdorp-Vogelaar I, et al. Cost-effectiveness Analysis of a Quantitative Immunochemical Test for Colorectal Cancer Screening. Gastroenterology. 2011;141(5):1648-55.

43. Gezondheidsraad. A National Colorectal Cancer Screening Programme. The Hague: Health Council of the Netherlands, 2008.

44. Boersma C, Broere A, Postma MJ. Quantification of the Potential Impact of Cost-effectiveness Thresholds on Dutch Drug Expenditures Using Retrospective Analysis. Value in Health. 2010;13(6):853-6.

45. Al MJ, Feenstra TL, van Hout BA. Optimal allocation of resources over health care programmes: dealing with decreasing marginal utility and uncertainty. Health Economics. 2005;14(7):655-67.

46. Bryan S, Williams I, McIver S. Seeing the NICE side of cost-effectiveness analysis: a qualitative investigation of the use of CEA in NICE technology appraisals. Health Economics. 2007;16(2):179-93.

47. Drummond M, Brown R, Fendrick AM, Fullerton P, Neumann P, Taylor R, et al. Use of Pharmacoeconomics Information: Report of the ISPOR Task Force on Use of Pharmacoeconomic/Health Economic Information in Health-Care Decision Making. Value in Health. 2003;6(4):407-16.

48. Drummond M, Cooke J, Walley T. Economic evaluation under managed competition: Evidence from the U.K. Social Science & Medicine. 1997;45(4):583-95.

49. Hoffmann C, Stoykova BA, Nixon J, Glanville JM, Misso K, Drummond MF. Do Health-Care Decision Makers Find Economic Evaluations Useful? The Findings of Focus Group Research in UK Health Authorities. Value in Health. 2002;5(2):71-8.

50. Hutton J, Brown RE. Use of Economic Evaluation in Decision Making: What Needs to Change? Value in Health. 2002;5(2):65-6.

51. Williams I, Bryan S. Understanding the limited impact of economic evaluation in health care resource allocation: A conceptual framework. Health Policy. 2007;80(1):135-43.

52. Coast J. Is Economic Evaluation in Touch with Society's Health Values? British Medical Journal. 2004;329(7476):1233-6.

53. Harris J. QALYfying the value of life. Journal of Medical Ethics. 1987;13(3):117-23.

54. Wald NJ, Oppenheimer P. Discounting the value of life. Journal of Medical Screening. 2011;18(1):1.

55. Sheldon TA. Discounting in health care decision-making: time for a change? Journal of Public Health. 1992;14(3):250-6.

56. Crott R. Economic analysis of HPV-vaccines: Not so simple? Vaccine. 2007;25(45):7717.

57. McGregor M, Caro JJ. QALYs: Are They Helpful to Decision Makers? PharmacoEconomics. 2006;24(10):947-52.

58. Neumann PJ. What next for QALYs? JAMA: The Journal of the American Medical Association. 2011;305(17):1806-7.

59. The Patient Protection and Affordable Care Act, (2010).

60. Caro JJ, Nord E, Siebert U, McGuire A, McGregor M, Henry D, et al. The efficiency frontier approach to economic evaluation of health-care interventions. Health Economics. 2010;19(10):1117-27.

61. Directions to Primary Care Trusts and NHS trusts in England concerning Arrangements for the Funding of Technology Appraisal Guidance from the National Institute for Clinical Excellence (NICE), (2003).

62. Neumann PJ. Why don't Americans use cost-effectiveness analysis? The American journal of managed care. 2004;10(5):308-12.

63. Neumann PJ, Greenberg D, Olchanski NV, Stone PW, Rosen AB. Growth and Quality of the Cost-Utility Literature, 1976-2001. Value in Health. 2005;8(1):3-9.

64. Drummond M, Sculpher M. Common Methodological Flaws in Economic Evaluations. Medical Care. 2005;43(7):II-5-II-14.

65. Drummond M. Cost-effectiveness guidelines for reimbursement of pharmaceuticals: Is economic evaluation ready for its enhanced status? Health Economics. 1992;1(2):85-92.

66. Sheldon TA. Problems of using modelling in the economic evaluation of health care. Health Economics. 1996;5(1):1-11.

67. Pinkerton SD, Johnson-Masotti AP, Derse A, Layde PM. Ethical issues in cost-effectiveness analysis. Evaluation and Program Planning. 2002;25(1):71-83.

68. Elsinga E, Rutten FFH. Economic evaluation in support of national health policy: The case of the Netherlands. Social Science and Medicine. 1997;45(4):605-20.

69. Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. Health Technology Assessment. 2008;12(7).

70. O'Mahony JF, de Kok IMCM, van Rosmalen J, Habbema JDF, Brouwer W, van Ballegooijen M. Practical implications of differential discounting in cost-effectiveness analyses with varying numbers of cohorts. Value in Health. 2011;14(4):438-42.

71. HM Treasury. Appraisal and Evaluation in Central Government (The Green Book). In: HM Treasury, editor. London2011.

72. Lazaro A. Theoretical arguments for the discounting of health consequences: Where do we go from here? PharmacoEconomics. 2002;20(14):943-61.

73. Weinstein MC, Stason WB. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. New England Journal of Medicine. 1977;296(13):716-21.

74. Frederick S, Loewenstein G, O'Donoghue T. Time Discounting and Time Preference: A Critical Review. Journal of Economic Literature. 2002;40(2):351-401.

75. Goodin RE. Discounting Discounting. Journal of Public Policy. 1982;2(01):53-71.

76. Brouwer WBF, van Exel NJA. Discounting in decision making: the consistency argument revisited empirically. Health Policy. 2004;67(2):187-94.

77. Harvey CM. The reasonableness of non-constant discounting. Journal of Public Economics. 1994;53(1):31-51.

78. Bleichrodt H, Gafni A. Time preference, the discounted utility model and health. Journal of Health Economics. 1996;15(1):49-66.

79. Lipscomb J, Weinstein MC, Torrence GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996. p. 214-46.

80. Cretin S. Cost/benefit analysis of treatment and prevention of myocardial infarction. Health services research. 1977;12(2):174-89.

81. West RR. Discounting the future: influence of the economic model. Journal of Epidemiology and Community Health. 1996;50(3):239-44.

82. Bonneux L, Birnie E. The discount rate in the economic evaluation of prevention: a thought experiment. Journal of Epidemiology and Community Health. 2001;55(2):123-5.

83. Lazaro A, Barberan R, Rubio E. The discounted utility model and social preferences: Some alternative formulations to conventional discounting. Journal of Economic Psychology. 2002;23(3):317-37.

84. Caro JJ, Huybrechts KF, Klittich WS. Secondary prevention fallacy: pitfalls in comparing with primary. Expert Review of Pharmacoeconomics & Outcomes Research. 2001;1(1):13-8.

85. Bos JM, Beutels P, Annemans LJP, Postma MJ. Valuing Prevention Through Economic Evaluation: Some Considerations Regarding the Choice of Discount Model for Health Effects with Focus on Infectious Diseases. PharmacoEconomics. 2004;22(18):1171-9.

86. Newhouse JP. Medical-Care Expenditure: A Cross National Survey. The Journal of Human Resources. 1977;12(1):115-25.

87. Hall RE, Jones CI. The Value of Life and the Rise in Health Spending. The Quarterly Journal of Health Economics. 2007;122(1):39-72.

88. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics. 2001;10(7):587-99.

89. Parsonage M, Neuburger H. Discounting and Health Benefits. Health Economics. 1992;1(1):71-6.

90. Viscusi WK. Discounting Health Effects for Medical Decisions. In: Sloan FA, editor. Valuing Health Care: Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies. Cambridge: Cambridge University Press; 1995. p. 125-47.

91. van Hout B. Discounting costs and effects: a reconsideration. Health Economics. 1998;7(7):581-94.

92. Klok RM, Brouwer WBF, Annemans LJP, Bos JM, Postma MJ. Towards a healthier discount procedure. Expert Review of Pharmacoeconomics and Outcomes Research. 2005;5(1):59-63.

93. Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal. 2005;331(7514):446-8.

94. Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: stepping forward towards optimal decision rules. Health Economics. 2007;16(3):307-17.

95. Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. Management Science. 1983;29(3):300-6.

96. Claxton K, Sculpher M, Culyer A, McCabe C, Briggs A, Akehurst R, et al. Discounting and cost-effectiveness in NICE - stepping back to sort out a confusion. Health Economics. 2006;15(1):1-4.

97. NICE. Discounting of Health Benefits in Special Circumstances. National Institute for Health and Clinical Excellence, 2011.

98. AOTM. Guidelines for conducting Health Technology Assessment (HTA). Warsaw: Agencja Oceny Technologii Medycznych, 2009.

99. Cleemput I, van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Brussels: Health Care Knowledge Centre (KCE), 2008 Contract No.: 78C.

100. Oostenbrink JB, Bouwmans CAM, Koopmanschap MA, Rutten FFH. Guidelines for Pharmacoeconomic Research, Updated Version. Diemen: College voor Zorgverzekeringen, 2004.

101. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health care technologies. Health Economics. 2011;20(1):2-15.

102. McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold: What it is and What that Means. PharmacoEconomics. 2008;26(9):733-44.

103. Culyer A, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. Journal of Health Services Research & Policy. 2007;12(1):56-8.

104. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What Is the Price of Life and Why Doesn't It Increase at the Rate of Inflation? Archives of Internal Medicine. 2003;163(14):1637-41.

105. Teljeur C, Flattery M, Harrington P, O'Neill M, Moran PS, Murphy L, et al. Cost-effectiveness of prion filtration of red blood cells to reduce the risk of transfusion-transmitted variant Creutzfeldt-Jakob disease in the Republic of Ireland. Transfusion. 2012;DOI: 10.1111/j.1537-2995.2012.03637.x.

106. O'Mahony JF. Differential Discounting: Questioning the Assumption of Healthcare Resource Fungibility over Time (Conference Poster) ISPOR 14th Annual European Congress; Madrid2011.

107. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. Health Economics. 2011;21(5):612-8.

108. Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Medical Decision Making. 2010;30(4):426-37.

109. Lipscomb J. Time Preference for Health in Cost-Effectiveness Analysis. Medical Care. 1989;27(3):S233-S53.

110. Dewilde S, Anderson R. The cost-effectiveness of screening programs using single and multiple birth cohort simulations: A comparison using a model of cervical cancer. Medical Decision Making. 2004;24(5):486-92.

111. Karnon J, Brennan A, Akehurst R. Decision Modeling to Inform Decision Making: Seeing the Wood for the Trees. Medical Decision Making. 2010;30(3):E20-E2.

112. Davies R, Roderick P, Raftery J. The evaluation of disease prevention and treatment using simulation models. European Journal of Operational Research. 2003;150(1):53-66.

113. Schulpher M. Subgroups and Heterogeneity in Cost-Effectiveness Analysis. PharmacoEconomics. 2008;26(9):799-806.

114. O'Mahony JF, van Rosmalen J, Zauber AG, van Ballegooijen M. Multicohort Models in Cost-Effectiveness Analysis: Why Aggregating Estimates over Multiple Cohorts Can Hide Useful Information. Medical Decision Making. 2012; DOI:10.1177/0272989X12453503.

115. Parmigiani G. Timing medical examinations via intensity functions. Biometrika. 1997;84(4):803-16.

116. Parmigiani G, Kamlet M. A Cost-Utility Analysis of Alternative Strategies In Screening For Breast Cancer. In: Gatsonis C, Hodges J, Kass RE, Singpurwalla N, editors. Bayesian Statistics in Science and Technology: Case Studies. I. New York: Springer-Verlag; 1993. p. 390-402.

117. de Gelder R, Bulliard JL, de Wolf C, Fracheboud J, Draisma G, Schopper D, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. European Journal of Cancer. 2009;45(1):127-38.

118. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JDF. Cost-effectiveness of cervical cancer screening: Comparison of screening policies. Journal of the National Cancer Institute. 2002;94(3):193-204.

119. Beutels P, Jit M. A brief history of economic evaluation for human papillomavirus vaccination policy. Sexual Health. 2010;7(3):352-8.

Paper 1.

Practical Implications of Differential Discounting in Cost-Effectiveness Analyses with Varying Numbers of Cohorts

James F. O'Mahony, MA[1,2],
Inge M. C. M. de Kok, MSc[2],
Joost van Rosmalen, PhD[2],
J. Dik F. Habbema, PhD[2],
Werner Brouwer, PhD[3],
Marjolein van Ballegooijen, MD, PhD[2].

1. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland;
2. Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands;
3. Institute of Health Policy and Management and Institute of Medical Technology and Assessment, Erasmus University Medical Centre, Rotterdam, the Netherlands.

Value in Health 14(4).

ABSTRACT

Objective: To call attention to the influence of the number of birth-cohorts used in cost-effectiveness analysis (CEA) models on incremental cost-effectiveness ratios (ICERs) under differential discounting.

Methods: The consequences of increasing the number of birth-cohorts are demonstrated using a CEA of cervical cancer prevention as an example. The cost-effectiveness of vaccinating 12-year-old girls against the Human Papillomavirus is estimated with the MISCAN model for 1, 10, 20 and 30 birth-cohorts. Costs and health effects are discounted with equal rates of 4% and alternatively with differential rates of 4% and 1.5% respectively. The effects of increasing the number of cohorts are shown by comparing the ICERs under equal and differential discounting.

Results: The ICER decreases as the number of cohorts increases under differential discounting, but not under equal discounting.

Conclusions: The variation of ICERs with the number of cohorts under differential discounting prompts questions regarding the appropriate specification of CEA models and interpretation of their results. In particular, it raises concerns that arbitrary variation in study specification leads to arbitrary variation in results. Such variations could lead to erroneous policy decisions. These findings are relevant to CEA guidance authorities, CEA practitioners and decision makers. Our results do not imply a problem with differential discounting per se, yet they highlight the need for practical guidance for its use.

INTRODUCTION

Debate over Differential Discounting

Discounting is used in cost-effectiveness analysis (CEA) to adjust future costs and health effects to their present values and volumes. This adjustment is to account for the positive time preference for goods, including health [1]. While discounting in general is widely accepted in CEA and other forms of economic analysis [2], whether discount rates for costs and effects should be equal has been debated extensively within health economics [2-9]. Equal discount rates for costs and effects are most commonly used [10]. Equal discounting is supported by a number of arguments, the most important of which are Weinstein and Stason's consistency thesis, Keeler and Cretin's postponing paradox and the tradability of health argument [3,4,11]. Increasingly however, differential discounting is advocated, whereby health effects are discounted at a different (typically lower) rate than costs [2,6,7,9,12]. Previous arguments for differential discounting were primarily based on the anticipation of an increasing societal value of health as income grows [7,12,13]. Recent work has shown, more generally, that differential discounting is justified if the cost-effectiveness threshold is anticipated to change, where the threshold may be defined with reference to either the consumption value of health or the cost-effectiveness of displaced interventions at the margin in the context of fixed health care budgets [14].

Currently, only a small number of CEA authorities recommend differential discounting [15]. The Dutch Health Care Insurance Board (College voor Zorgverzekeringen [CVZ]) revised its recommended rates in 2006, from equal rates of 4% to differential rates of 4% and 1.5% for costs and health effects respectively [12,16]. Belgium also recently adopted differential discounting at rates of 3% and 1.5% for costs and effects respectively [17]. The National Institute for Health and Clinical Excellence (NICE) in England and Wales used differential discounting from its inception with rates of 6% and 1.5% for costs and effects respectively, but reverted to equal discounting at 3.5% in 2004 [7,18].

Modelling and Decision Rules in Cost-Effectiveness Analysis

Modelling is widely used in cost-effectiveness analysis of healthcare interventions [19]. Modelling can be used both to extrapolate outcomes beyond trial follow-up periods and to simulate interventions that have not or cannot be assessed using controlled trials [1,20,21]. CEA models most commonly only simulate one cohort of individuals [22]. A multiple-cohort modelling approach is more appropriate in some cases, such as where risk factors change over time, leading to cohort effects; where the effects of a disease are dynamic, such as in infectious diseases; or, where both prevalent and incident cohorts need to be considered [22-25].

Decision making using CEA relies on comparisons between analyses to determine which interventions are cost-effective. In theory, interventions can be ranked by their incremental cost-effectiveness ratios (ICERs) and accepted in order of cost-effectiveness until the budget constraint is reached [26]. In practice, it is more typical to accept interventions with ICERs below a given threshold as cost-effective [4]. Both decision rules compare interventions' ICERs, either directly in the case of the ranking rule, or indirectly through the threshold.

Overview

We compare the results of single-cohort and multiple-cohort CEAs of the same intervention to quantify the consequences of alternative numbers of cohorts (henceforth CEA specification) under differential discounting. We show, using a CEA of vaccination against the Human Papillomavirus (HPV) as an illustrative example, that the ICER falls as more cohorts are included in the analysis under differential discounting, but remains constant under equal discounting. Recent work by Hoyle and Anderson also notes that increasing the number of cohorts reduces ICERS under differential discounting [22].We address this particular issue in greater depth and consider its significance for CEA practice and health-care decision making. In this article, we take no normative stance for or against differential discounting; however, we consider its consequences from the perspective of equal discounting being the policy norm in most countries to date. Most previous studies of differential discounting have addressed its theoretical merits; our study adds to the literature by explicitly considering the practical consequences of differential discounting for decision making using CEA.

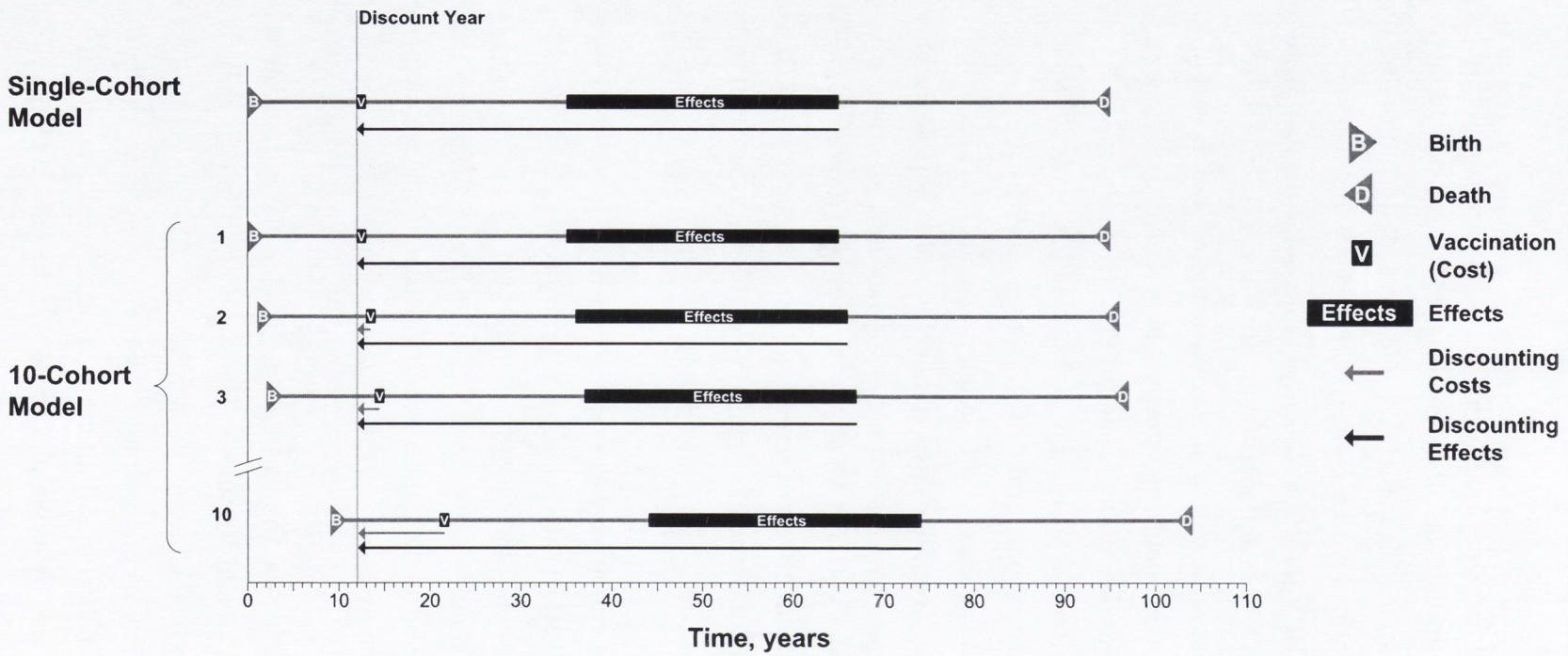
METHODS

We examine how an intervention's ICER changes as the number of cohorts in a CEA model increases. We use the example of the MISCAN microsimulation screening analysis model of HPV vaccination in the Netherlands used in a recently published CEA [27]. Further detail of the model specification and assumptions can be found in that publication. The model simulates the individual life histories of one or more birth-year cohorts of women, who, in the absence of either screening or vaccination, acquire a HPV infection at a certain rate, some develop a pre-invasive lesion and/or cancer, of whom a proportion die from the disease. This results in an age and calendar time-specific output of disease incidence and mortality. The simulation is repeated, now including screening both with and without vaccination. Screening is simulated as the current (as of 2011) Dutch programme: 7 screens between the ages of 30-60 at 5 year intervals. Vaccination is administered at age 12. These interventions change some of the life histories, either by preventing disease or detecting and treating it earlier, resulting in improved health states, longer life and reduced treatment costs. Treatment costs and quality of life adjustments are then applied to these consequences to estimate the intervention's treatment cost savings and quality-adjusted life year (QALY) gains. The difference between the total net discounted costs and health effects of screening alone and screening and vaccination combined is used to calculate the incremental cost-effectiveness of HPV vaccination.

A number of additional simplifying assumptions are made in the present study: the undiscounted costs and effects are the same for every cohort; no booster vaccination is required; and, there are no start-up or fixed programme costs. A large number of women (1 billion) are simulated in each model to minimise differences due to random error. Each cohort in the multi-cohort models contains an equal proportion of the total number of individuals.

The ICER of HPV vaccination from a single-cohort analysis is compared to ICERs from analyses with 10, 20 and 30 cohorts. The analyses differ only in the number of cohorts used; all else is held constant. Each cohort is defined by its birth year and each receives the vaccination one year after the preceding cohort. Figure 1 depicts a single-cohort and a 10-cohort model. Costs and effects are discounted by 4% and 1.5% respectively and also by a common rate of 4%. Costs and effects are discounted to the year the first cohort is vaccinated.

Figure 1: Comparison of discounting in a single cohort and a multi-cohort model with 10 cohorts (cohorts 4–9 not shown).



RESULTS

Table 1 presents the results of the single and multi-cohort models under equal and differential discounting. The table reports the discounted incremental costs and effects and the corresponding ICERs in each of the models and discount rate assumptions. The table also reports the ICERs of the multi-cohort models as a percentage of that of the single-cohort model.

The ICERs are significantly lower under differential discounting compared to equal discounting. This is due to the lower discounting of health effects; consequently the discounted effects are larger while the discounted costs remain the same, resulting in lower ICERs.

The important result, however, is the variation in ICERs between models with different numbers of cohorts. The ICERs do not vary significantly with the number of cohorts under equal discounting; the small differences in costs and effects are due to random variation in the simulation model. Conversely, with differential discounting the ICERs are considerably lower in the multi-cohort models and fall as the number of cohorts increases. The magnitude of the differences is large: the 10-cohort model has an ICER that is approximately 90% of the single-cohort model's ratio; the 30-cohort model has an ICER that is approximately 74% of the single-cohort model's ratio.

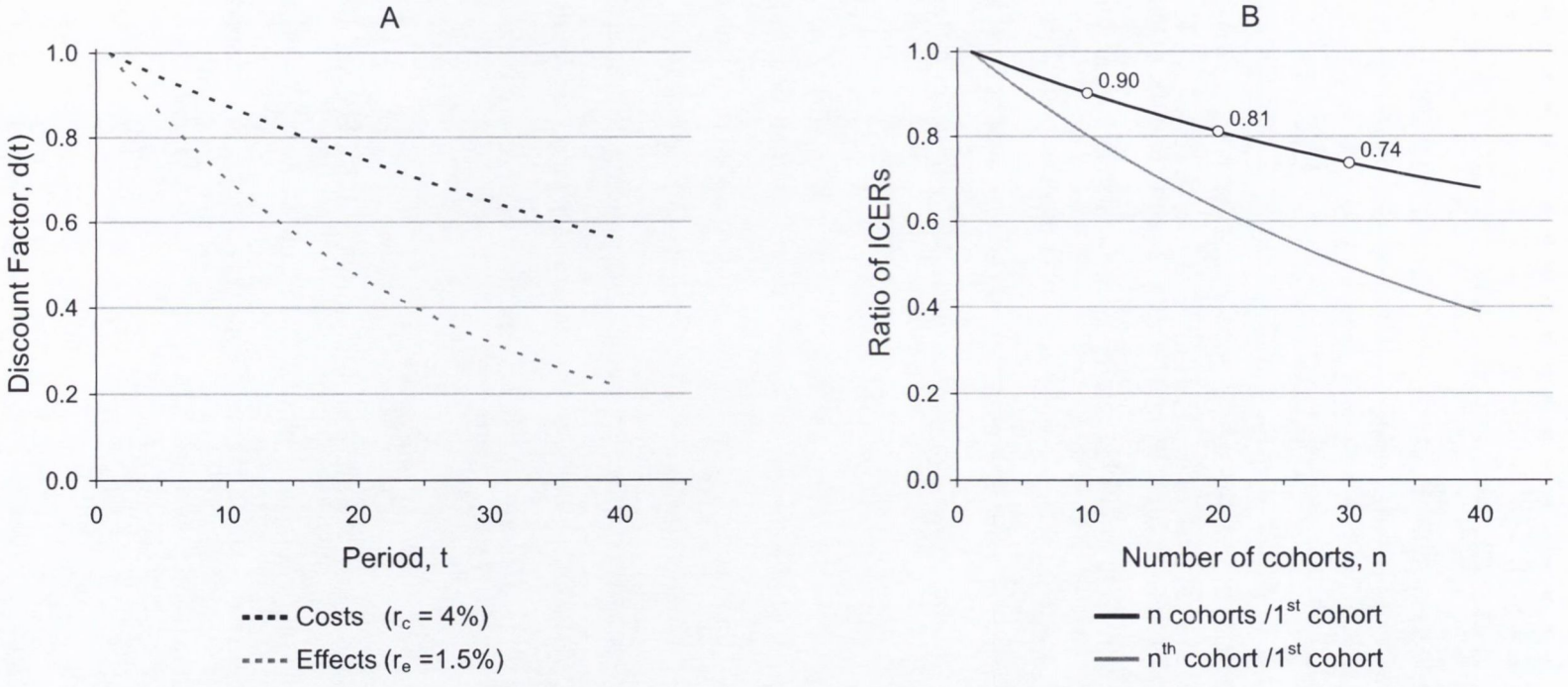
Panel A of Figure 2 shows the annual discount factors for the current Dutch discount rates for cost and effects of 4% and 1.5 % respectively, where t=1 is the discount year. The grey line in panel B shows the ratios of ICERs of the n^{th} future cohort relative to the 1^{st} cohort at the discount year, where each cohort is n-1 years from the discount year and undiscounted costs and effects are equal for all cohorts. This line is also the ratio of the discount factors shown in panel A. The black line in panel B represents the ratio of ICERs of a multi-cohort model with n cohorts relative to a model of a single cohort at the discount year, with the points labelled for models with n = 10, 20 and 30 cohorts.

Table 1: Incremental costs, incremental health effects and ICER of adding vaccination against the Human Papillomavirus 16/18 to the current screening programme in the Netherlands, a comparison of a single cohort and multiple cohort models with 10, 20 and 30 cohorts under equal discounting of 4% for costs and effects and differential discounting of 4% and 1.5% for costs and effects respectively

	Equal Discount rates: 4% & 4%				Differential discount rates: 4% & 1.5%			
	Single Cohort	10 Cohorts	20 Cohorts	30 Cohorts	Single Cohort	10 Cohorts	20 Cohorts	30 Cohorts
Incremental costs, €M	324,423	273,662	229,268	194,470	324,423	273,662	229,268	194,470
Incremental effects, QALYs (000s)	3,190	2,690	2,254	1,912	10,839	10,146	9,444	8,809
ICER, €/QALY (4 s.f.)	101,700	101,700	101,700	101,700	29,900	27,000	24,300	22,100
Ratio of ICERs, multiple/single cohort	reference	100%	100%	100%	reference	90%	81%	74%

QALY=quality-adjusted life-year; ICER= incremental cost-effectiveness ratio:

Figure 2: (A) Annual discount factors over time under discount rates of 4% and 1.5% for costs and effects respectively; (B) ratio of ICERs of the nth single cohort relative to the 1st single cohort at the discount year and the ratio of ICERs of a multi cohort model with n cohorts relative to the 1st single cohort under simplifying assumptions.



DISCUSSION

Explanation

Our results are easily explained by considering the differences between equal and differential discounting. Under equal discounting, varying the length of time between the discount year and the intervention (and its effects) does not influence the ICER, because although the present value of costs and effects change, they vary proportionately. It is in this respect that Lipscomb et al. describe CEA as "time neutral" under equal discounting [28]. CEA is not "time neutral" under differential discounting. Increasing the length of time between the intervention and the discount year causes the present value of effects to fall less than proportionally to the reduction in the present value of costs, resulting in a lower ICER. Consequently, both shifting a single cohort to a later period relative to the discount year and adding later cohorts to a CEA will not cause the ICER to fall under equal discounting, but will under differential discounting. Our analysis demonstrates the second of these two effects.

We have highlighted the consequences of a lower discount rate on health effects, which is appropriate if the threshold is growing, as the discount differential should approximate the annual growth rate of the threshold [14]. However, the threshold may not necessarily grow over time, even with an expanding health care budget, but may be static or fall [29]. A falling threshold would imply a larger discount rate for health than costs [30], resulting in increasing ICERs as more cohorts are included.

Relevance for Practice and Policy

The analysis shows how the number of cohorts used in CEA can influence ICERs. To understand the practical significance of this result we have to consider current CEA practice. The current understanding of appropriate CEA model specification most likely does not account for the influence of varying numbers of cohorts. Consequently, without clear guidance on the matter, CEA practitioners are likely to continue specifying studies with the minimum number of cohorts they consider necessary; because interventions differ in their modelling requirements, this will continue to result

in the variety of models specifications evident in reviews of modelling methodologies [24,31,32].

The concern is that arbitrary variation in CEA specifications leads to results which, in part, vary arbitrarily too. A related concern is that CEAs may be deliberately specified with large numbers of cohorts or large lags between the discount year and the start of the intervention to achieve low ICERs. We have focused on the issue of multiple cohorts rather than unnecessarily long lags between the discount year and implementation because the latter is more easily recognisable as inappropriate manipulation of the CEA. Such arbitrary or strategically chosen variations can compromise comparability between studies. As a result CEAs, may not be adequately reflecting the policy choices they are intended to inform. These concerns are compounded by the probable lack of awareness among decision makers of the influence of CEA specifications on results. Decision makers may well continue comparing ICERs directly, without taking the different model specifications into account or checking whether they adequately reflect the relevant policy choice. Such direct comparisons could lead to incorrect policy choices, whereby an intervention is deemed cost-effective as it has a lower ICER than the threshold or an alternative intervention, but where this result is due to arbitrarily or strategically chosen differences in the CEA specification, rather than the intervention's actual implementation and inherent characteristics.

Naturally, these concerns lead to a consideration of how to avoid or reduce arbitrary variations between studies. For instance, one could prescribe a base-case specification that imposes a standard number of cohorts for all CEAs. However, given the wide variation of both the characteristics and implementation of interventions, a standardised CEA specification may not adequately reflect these differences and thus result in meaningless comparisons. Therefore, if standardisation is not possible, it is not yet clear how or if CEA practice can be adapted to avoid arbitrary variation between studies. Consequently, CEAs should be evaluated on an informed, case by case basis. Accordingly, it is appropriate to demand a clear justification of the CEA specification from the CEA practitioner.

These questions of how to specify and interpret CEAs relate to doubts about the appropriateness of current CEA decision rules. A number of authors have indicated that ICERs are inappropriate for determining the optimal timing of interventions [2,33-35]. Counter-intuitive results arise under both equal and differential discounting. For example, decision makers choosing the period of implementation with the lowest ICER will prefer to (infinitely) postpone implementation under differential discounting (the postponing paradox) [3], whereas, unrealistically, they should be indifferent between immediate implementation or infinite postponement under equal discounting. While both these results are difficult to defend, it is the postponing paradox that has generated debate in the literature. The postponing paradox has been dismissed as irrelevant to actual policy choices [6]. Indeed, when using a threshold based decision rule, any postponement will not be infinite, but until the ICER falls below the threshold.

While the relevance of the postponing paradox to actual policy choices is disputed, our results show that interventions modelled with a greater proportion of their implementation in the future are advantaged by lower ICERs. In this context, Cohen's questioning of the appropriateness of current decision rules to health care services that exhaust their budgets annually without saving a surplus may be relevant [34]. He commented that using CEA to compare interventions over multiple periods implies that cohorts compete for resources that are fungible across periods, whereas it might be more appropriate to use CEA to compare cohorts competing for resources within periods. Consequently, the debate over differential discounting and the implications for comparisons between studies may prompt a broader reconsideration of policy decision rules and the economic evaluation of healthcare.

Recommendations for CEA Practice

The aim of this study is to promote awareness of the effects of alternative CEA specifications under differential discounting among CEA practitioners and decision makers. We hope CEA advisory bodies will recognise the significance of the findings presented here and reflect it in their guidance, for example:1) by requiring a justification of the CEA's specification; and 2) by providing guidance to decision makers regarding the influence of the number of cohorts included. Such clarity is important, as confusion regarding the validity of comparisons between analyses can only serve to damage CEA's credibility with decision makers and others. Note that the issues raised in this paper should not be interpreted as arguments against differential discounting; rather they should be understood as a call for greater understanding of its practical implications. CEA authorities considering adopting differential discounting should consider these practical implications in addition to the theoretical arguments. CEA practitioners and decision makers in countries already using differential discounting would benefit from recognising its implications to ensure best practice and correct policy choices.

Limitations and Further Research

We emphasise that this paper does not provide a complete discussion of the methodological implications of differential discounting. Such a discussion would require a detailed review of the underlying theoretical basis for comparing interventions across different time periods with a changing threshold, which is beyond the scope of this study. This remains an important area for future research and debate; see Claxton et al.[14] for further discussion. However, this paper does call attention to some important practical issues related to differential discounting that both analysts and policy makers need to be aware of when using and comparing the results of cost-effectiveness analyses in practice.

REFERENCES

1. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.

2. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics 2001;10(7):587-99.

3. Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. Management Science 1983;29(3):300-6.

4. Weinstein MC, Stason W. Foundations of cost-effectiveness analysis for health and medical practices. New England Journal of Medicine 1977;296(13):716-21.

5. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996.

6. van Hout B. Discounting costs and effects: a reconsideration. Health Economics 1998;7(7):581-94.

7. Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal 2005;331(7514):446-8.

8. Claxton K, Sculpher M, Culyer A, McCabe C, Briggs A, Akehurst R, et al. Discounting and cost-effectiveness in NICE - stepping back to sort out a confusion. Health Economics 2006;15(1):1-4.

9. Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: stepping forward towards optimal decision rules. Health Economics 2007;16(3):307-17.

10. Smith DH, Gravelle H. The practice of discounting in economic evaluations of healthcare interventions. International Journal of Technology Assessment in Health Care 2001;17(2):236-43.

11. Williams A. Welfare economics and health status management. In: Health, Economics and Health Economics, 1981; Amsterdam: p. 123-32.

12. Klok RM, Brouwer WBF, Annemans LJP, Bos JM, Postma MJ. Towards a healthier discount procedure. Expert Review of Pharmacoeconomics and Outcomes Research 2005;5(1):59-63.

13. Meerding WJ, Bonsel G, J, Brouwer W, B. F, Stuifbergen M, C, Essink-Bot M-L. Social time preferences for health and money elicited with a choice experiment. Value in Health 2010;13(4):368–74.

14. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health care technologies. Health Economics 2011;20(1):2-15.

15. ISPOR. Pharmacoeconomic guidelines around the world [cited 2009 10-6-09]; Available from: http://www.ispor.org/PEguidelines/index.asp.

16. College voor Zorgverzekeringen. Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie. 2006.

17. Federaal Kenniscentrum voor de Gezondheidszorg. Guidelines for pharmacoeconomic evaluations in Belgium. Brussels; 2008.

18. Severens J, Milne R. Discounting health outcomes in economic evaluation: the ongoing debate. Value in Health 2004;7(4):397-401.

19. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of Health Services Research and Policy 2004;9(2):110-8.

20. Buxton M, J. , Drummond M, F. , van Hout B, A. , Prince R, L. , Sheldon T, A. , Szucs T, et al. Modelling in Economic Evaluation: An Unavoidable Fact of Life. Health Economics 1997;6(3):217-27.

21. Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2001.

22. Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Medical Decision Making 2010;30(4):426-37.

23. Dewilde S, Anderson R. The cost-effectiveness of screening programs using single and multiple birth cohort simulations: A comparison using a model of cervical cancer. Medical Decision Making 2004;24(5):486-92.

24. Kim JJ, Brisson M, Edmunds WJ, Goldie SJ. Modeling cervical cancer prevention in developed countries. Vaccine 2008;26(S10):K76-K86.

25. Habbema JDF, Boer R, Barendregt J. Chronic Disease Modeling. In: Heggenhougen K, Quah S, editors. International Encyclopedia of Public Health. San Diego: Academic Press; 2008. p. pp. 704-9.

26. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 2004;13(5):437-52.

27. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of human papillomavirus vaccination in the Netherlands. Journal of the National Cancer Institute 2009;101(15):1083-92.

28. Lipscomb J, Weinstein MC, Torrence GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996. p. 214-46.

29. McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold: What it is and What that Means. Pharmacoeconomics 2008;26(9):733-44.

30. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. Health Economics 2011;21(5):612-8.

31. Beutels P, Van Doorslaer E, Van Damme P, Hall J. Methodological issues and new developments in the economic evaluation of vaccines. Expert Review of Vaccines 2003;2(5):649-60.

32. Cooper K, Brailsford S, Davies R, Raftery J. A review of health care models for coronary heart disease interventions. Health Care Management Science 2006;9(4):311-24.

33. Cairns J. Discounting and health benefits: another perspective. Health Economics 1992;1(1):76-9.

34. Cohen BJ. Discounting in cost-utility analysis of healthcare interventions: Reassessing current practice. Pharmacoeconomics 2003;21(2):75-87.

35. Parsonage M, Neuburger N. Discounting and health benefits. Health Economics 1992;1(1):71-6.

Paper 2.

Correcting for Multiple Future Cohorts when Applying Differential Discounting of Costs and Health Effects

James F. O'Mahony MA[1,2],
Maarten J. Postma PhD[3,4],
Cornelis Boersma PhD[3],
Tjalke A. Westra MSc[5],
Joost van Rosmalen PhD[6,2],
Mark Jit PhD[7,8].

1. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland.
2. Department of Public Health, Erasmus Medical Centre, Rotterdam, the Netherlands.
3. Unit of PharmacoEpidemiology and PharmacoEconomics, Department of Pharmacy, University of Groningen, Groningen, the Netherlands.
4. Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands.
5. Department of Medical Microbiology, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands.
6. Department of Biostatistics, Erasmus Medical Centre, Rotterdam, the Netherlands.
7. Health Protection Agency, London, UK.
8. Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK.

Under revision following submission.

ABSTRACT

Background: Differential discounting of costs and health effects means cost-effectiveness analyses with more future cohorts will give more favourable cost-effectiveness estimates, all else equal. This prevents direct comparisons between cost-effectiveness analyses with different numbers of cohorts.

Objective: To show how to make fair comparisons of cost-effectiveness between models with different numbers of future cohorts under differential discounting.

Methods: The cost-effectiveness estimate of vaccination against human papillomavirus (HPV) infection from a published cost-effectiveness analysis simulating 101 recipient cohorts over 101 years is compared to that of a hypothetical intervention with cost-effectiveness equal to the current cost-effectiveness threshold, assumed to be £20,000/QALY. The cost-effectiveness of the hypothetical intervention is considered when implemented with the same number of future cohorts as the HPV analysis. Implemented in this way, the cost-effectiveness of the hypothetical intervention serves as an adjusted threshold that accounts for the number of future recipient cohorts. An alternative adjustment is also described, whereby rather than adjusting the threshold, we adjust the cost-effectiveness estimate of the intervention itself for the number of future cohorts. This is achieved by dividing the intervention's cost-effectiveness estimate by the ratio of the adjusted threshold to the current threshold.

Results: The resulting adjusted cost-effectiveness estimate permits direct comparisons to other interventions and the current threshold. Adjusting for the number of cohorts results in a less favourable cost-effectiveness estimate of HPV vaccination, but it remains highly cost-effective.

Conclusions: A simple correction for the number of future recipient cohorts provides a fairer basis for comparison between interventions when differential discounting is used.

INTRODUCTION

Background to Differential Discounting

Economic growth is anticipated to lead to growth in the willingness-to-pay for health [1]. Differential discounting of costs and health effects in CEA is one way to account for this growth in willingness-to-pay. Typically, differential discounting involves applying a lower discount rate to health effects than costs.

Differential discounting has prompted robust debate [1-5]: however, a degree of consensus was achieved in a recent paper by Claxton et al. [6]. It shows that the appropriate form of differential discounting depends on the decision rules used and whether the maximand is health or welfare. Claxton et al. show that if a cost-effectiveness threshold is used to determine reimbursement eligibility, then the differential between the discount rates on costs and health effects should approximate the annual rate of threshold growth.

Differential discounting is required by CEA guidelines in the Netherlands (4 & 1.5% for costs and health effects respectively), Belgium (3 & 1.5%) and Poland (5 & 3.5%) [7-9]. Differential discounting was required by the National Institute for Health and Clinical Excellence (NICE) prior to 2004 (6 & 1.5%) [2]. NICE currently requires equal discounting at 3.5%: however, where an intervention is substantial in restoring health and provides benefits sustained over 30 years or more, NICE recommends differential rates of 3.5 & 1.5% [10]. Similarly, guidelines from NICE and in many other countries require differential discounting as part of a sensitivity analysis. Furthermore, differential discounting is sometimes applied where it is not required by guidelines [11-13].

The Problem of Comparability

One theoretical argument against differential discounting is the postponing paradox [14]. Differential discounting means an intervention will be relatively more cost-effective when implemented in next year's cohort rather than the current cohort. The paradox is that decision makers seeking the best cost-effectiveness will perpetually postpone implementation [15]. The paradox has been dismissed as irrelevant to decision making practice on the contention that cost-effectiveness estimates guide the choice of what to implement, rather than the timing of implementation [16].

O'Mahony et al. identify a problem of comparability between CEAs related to the postponing paradox [17]. They show that interventions have lower cost-effectiveness ratios when modelled with more future cohorts under differential discounting, all else equal. Consequently, it is difficult to compare interventions when the CEAs used to estimate cost-effectiveness feature different numbers of future cohorts. For example, if intervention A has a lower cost-effectiveness ratio than B, but is modelled with many future cohorts while B is modelled only in one current cohort, then it is unclear whether A's apparent superior cost-effectiveness is because it produces more health at less cost or because it was modelled in more cohorts.

Differential discounting in multiple cohort models also confuses comparisons to the current cost-effectiveness or willingness-to-pay threshold. An intervention implemented in a current cohort may not be cost-effective compared to the current threshold. However, adding multiple future cohorts to the analysis can yield a cost-effectiveness ratio under the current threshold. Consequently, the same intervention can appear both cost-effective and not cost-effective relative to the current threshold, depending on the number of future cohorts modelled.

To further understand the comparability problem it is helpful to distinguish between two effects of differential discounting. Within a given cohort, applying a lower discount rate to effects improves the cost-effectiveness of any intervention yielding effects after the intervention's costs are incurred. We describe this as the intra-cohort effect. A lower rate on effects also means the cost-effectiveness of any intervention improves if implemented in periods later than the discount year (the

discount year is synonymous with the baseline year): i.e. in cohorts starting the intervention after the discount year. We describe this as the inter-cohort effect.

The policy relevance of comparability is illustrated by a recent debate over the correct application of discounting. Comparability concerns prompted de Kok et al. to apply equal discounting when analysing HPV vaccination in the Netherlands [18, 19]. Consequently, their cost-effectiveness estimate was markedly less favourable than two other Dutch analyses that used differential discounting [20, 21]. This prompted criticism that de Kok et al. had not adhered to Dutch CEA guidelines [21, 22], and led to a misleading conclusion that the vaccine was not cost-effective [23].

Possible Solutions

Further context to our solution to the comparability problem is provided by two other potential solutions. One is to standardise the number of cohorts. The second is to use a two-step discounting procedure. However, neither are practical solutions for the reasons explained below.

Many CEAs only consider an intervention's current recipient cohort, despite the fact that there will be other cohorts in subsequent years [24]. This may reflect the implicit assumptions that cost-effectiveness will not change between cohorts and that the threshold will not fall over time. Given these assumptions, simulating the current cohort is sufficient to determine cost-effectiveness, both now and in the future.

There are, however, cases in which including future cohorts is necessary. Interventions against infectious diseases yield direct and indirect effects due to herd immunity. Indirect effects are shared across cohorts and can take years to emerge. Modelling the current cohort alone underestimates cost-effectiveness, as it ignores indirect effects that emerge in later cohorts [25, 26]. Furthermore, static single cohort models can fail to capture shifts in the age of infection that can occur in infectious disease control [25].

Another example in which modelling future cohorts may be necessary is the case of capital equipment, such as diagnostic imaging machines. As such capital can remain

in use for many years there will be multiple cohorts using the equipment. Since not all medical capital will have the same lifespan, the number of recipient cohorts will vary. For example, cobalt-60 radiotherapy units typically last 5-6 years, whereas linear accelerator units may last 10-12 years [27]. Consequently, standardising the number of cohorts modelled will be incompatible with the equipment's actual use.

A further example of where a multi-cohort model may be appropriate is where a broader change in health service provision occurs that affects multiple consecutive cohorts over a number of years. An example of such a change could be the shift in provision of elective procedures from an inpatient to an ambulatory service basis. The effects of such changes may occur progressively over a number of years, consequently, appraisal of such a change in provision in any one year would be misleading, so a model accounting for multiple cohorts over multiple years would be more appropriate.

An alternative to standardising the number of cohorts that would avoid the problem of comparability would be to employ two-step discounting. Two-step discounting has previously been suggested to reconcile differences between private and social discount rates [28-30]. In the case of differential discounting, the first step is to differentially discount costs and effects to the year each cohort starts the intervention; the second step is to equally discount the present values of costs and effects from each cohort's start year to a common discount year. The first step of differentially discounting within each cohort retains the intra-cohort effect of discounting, while equal discounting in the second step eliminates the inter-cohort effect.

While two-step discounting would solve the comparability problem, the theoretical rationale for its adoption is not immediately apparent. Furthermore, it is unclear what the common discount rate for the second step should be.

There is, however, a more fundamental problem with two-step discounting. In case of infectious disease control in which multiple cohorts receive an intervention over multiple years and there are shared indirect effects, these indirect effects, either health effects or cost savings, cannot be attributed to implementation of the

intervention in any particular cohort. Consequently, there is no identifiable discount year for the first step of discounting for these shared benefits.

The impracticality of standardising the number of cohorts and two-step discounting means a more general solution to the comparability problem is needed.

Purpose

The purpose of this paper is to present a method for adjusting cost-effectiveness estimates from multiple cohort models under differential discounting that preserves comparability between studies and to the current threshold. By resolving the comparability problem, our proposed adjustment will provide confidence in differential discounting and help avoid confusion, such as that exemplified by the debate over de Kok et al.'s results. Our analysis is intended to inform three groups: CEA analysts on how to perform our adjustment; decision makers on the interpretation of differentially discounted cost-effectiveness estimates; and, those drafting CEA guidelines.

METHODS

Model Background

Our proposed adjustment is illustrated using the example of a multi-cohort CEA of HPV vaccination developed by the UK's Health Protection Agency (HPA). The model description can be found in previous publications [31, 32]. The results used here correspond to the version of the model used in a recently published CEA comparing bivalent and quadrivalent vaccines under various scenarios by Jit et al. [31]. Specifically, the results correspond to the quadrivalent vaccine under scenario 7 considered in Jit et al., where the vaccine is assumed to have an average duration of protection of 20 years and to prevent all disease due to HPV subtypes 6,11,16 and 18. The cost-effectiveness of vaccination is estimated as the incremental cost-effectiveness ratio (ICER) of adding vaccination to the current UK cervical screening programme: i.e. the ICER is derived from a comparison of screening-only to vaccination-plus-screening, assuming screening remains unchanged.

The HPA model is a dynamic transmission model, capable of simulating both the direct and indirect effects of vaccination. The simulated indirect effects take over 60 years to approximate stability. Consequently, the model uses a 101 year time horizon, featuring 101 recipient cohorts: one recipient cohort in the baseline year and one in each subsequent year. All cost and effects are discounted to the baseline year when vaccination is first introduced.

The model simulates the UK's vaccination programme, which included catch-up vaccination during the first three years. Consequently, the first three cohorts constitute girls of mixed ages from 12 to 18 years and are larger than subsequent cohorts. Thereafter the vaccinated cohorts are of equal size and only include girls aged 12 to 13.

Jit et al. follow NICE's guidelines on discounting, using an equal rate of 3.5%, but also report a sensitivity analysis using differential rates of 3.5 & 1.5%. Consequently, Jit et al. provides an example of a CEA featuring multiple future cohorts and differential discounting.

Proposed Adjustment

NICE's stated threshold range is £20,000 – 30,000/QALY [33]. For convenience, we assume a point threshold of £20,000/QALY. This is the threshold in the discount year when the reimbursement decision will be made; henceforth described as the current threshold.

To correct for the effects of multiple future cohorts under differential discounting we consider a hypothetical intervention as a comparator to vaccination. We assume this hypothetical intervention is marginally cost-effective in the current period: i.e. its cost-effectiveness equals the current threshold. We assume it costs £20,000 in the year it is implemented and provides 1 QALY in the same year. These costs and effects are constant over time. No further direct or indirect health gains, costs or savings are assumed.

We consider the cost-effectiveness of this hypothetical intervention when implemented in the same number of cohorts as vaccination. We not only assume the

hypothetical intervention is also implemented over 101 years, but also that the distribution of spending over time on the hypothetical intervention is proportionate to that on vaccination. This is to match spending on the initially larger cohorts of the catch-up programme. When the hypothetical intervention is matched in this way, we describe it as the "hypothetical comparator".

The present value of the hypothetical comparator's costs and effects in the discount year yields what we describe as the "adjusted threshold", i.e. the current threshold adjusted for the number of future cohorts. If the ICER of vaccination is below this adjusted threshold then it is considered cost-effective; otherwise, we conclude vaccination is not cost-effective.

Figure 1 is a simplified representation of matching the hypothetical intervention to vaccination in three consecutive birth cohorts. The figure is not drawn to scale. It shows the hypothetical comparator's costs and effects occurring in the year of vaccination for each cohort. The figure also shows discounting of costs and effects for both vaccination and the hypothetical comparator to the discount year.

Figure 1: Simplified representation of the hypothetical intervention match to three cohorts of vaccination illustrating differential discounting

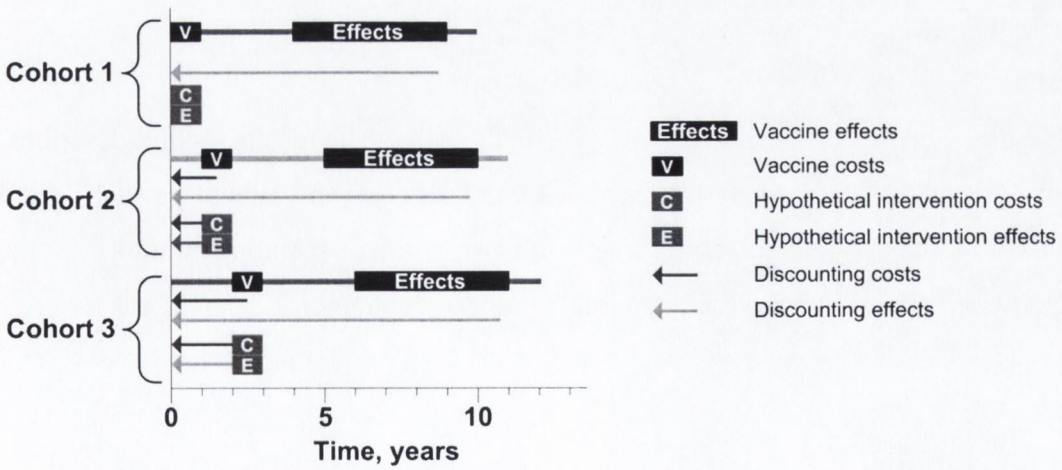


Table 2 details the hypothetical intervention matched to vaccination in the HPA model. The second column shows the undiscounted spend on vaccination in each cohort. The third column shows the matched spend on the hypothetical intervention

in each period. The fourth column shows the undiscounted effects of the hypothetical intervention, which are simply the cost of the hypothetical intervention divided by its cost-effectiveness of €20,000/QALY. The final two columns show the hypothetical intervention's discounted costs and effects, where the δ terms are the discount factors for costs and effects (where $\delta = 1/(1+r)$ and r is the discount rate). The ratio of the hypothetical comparator's total discounted costs and effects, shown in the last column's final row, gives the adjusted threshold.

This adjustment can be taken further to facilitate direct comparisons between interventions modelled with different numbers of future cohorts. Dividing the adjusted threshold by the current threshold gives what we describe as the "adjustment factor".

Dividing an intervention's estimated ICER by the adjustment factor yields what we describe as the "adjusted ICER". This adjusted ICER will have the same ratio to the current threshold as the intervention's unadjusted ICER to the adjusted threshold. Consequently, rather than adjusting the threshold for the number of future cohorts, we maintain the relative relationship between the ICER and the threshold by adjusting the ICER instead. This adjusted ICER can be compared directly to the current threshold to determine cost-effectiveness. Similarly, the adjusted ICERs of two interventions can be compared directly to determine their relative cost-effectiveness.

The adjustment factor can also be shown analytically. The results section describes the adjustment factor given the discount factors for costs and effects for each period, an expenditure weight corresponding to the proportion of spending on the intervention in each cohort and the total number of cohorts.

RESULTS

Table 1 reports the total incremental costs, health effects and cost-effectiveness ratio of vaccination over the 101 years of the HPA model, undiscounted, equally discounted at 3.5% and differentially discounted at 3.5 & 1.5%. It shows the ICER of vaccination under equal discounting at 3.5% to be £18,200/QALY. Vaccination is much more cost-effective under differential discounting, with an ICER of £7,400/QALY. The inter-cohort effect of differential discounting is so pronounced that the ICER under differential discounting is lower than under no discounting.

Table 1: Unadjusted cost-effectiveness of HPV vaccination under alternative discount rates

	Discount Assumptions		
	0 & 0%	3.5 & 3.5%	3.5 & 1.5%
ΔCosts, £M	7,637	2,552	2,552
ΔQALYs, (000s)	788	140	344
ICER, £/QALY (3s.f.)	9,700	18,200	7,400

Δ: Change in
QALY: Quality adjusted life year
ICER: Incremental cost-effectiveness ratio

Table 2 shows that the hypothetical comparator matched to vaccination yields an adjusted threshold of £11,200/QALY. Since the unadjusted ICER of vaccination under differential discounting of £7,400/QALY is below this adjusted threshold we conclude HPV vaccination is cost-effective, even when accounting for multiple future cohorts.

Dividing the adjusted threshold of £11,200/QALY by the current threshold yields an adjustment factor of 0.56. Dividing the vaccine's unadjusted ICER of £7,400/QALY by this adjustment factor yields an adjusted ICER of £13,300. Comparing this adjusted ICER to the current threshold of £20,000/QALY again indicates that vaccination is cost-effective. While the adjusted ICER is greater than the unadjusted ICER, it is still below the ICER of £18,200/QALY under equal discounting.

Table 2: Calculation of the adjusted cost-effectiveness threshold from the hypothetical comparator

Cohort	Vaccination Costs - undiscounted, £M	Hypothetical Comparator Costs - undiscounted, £M	Hypothetical Comparator Effects undiscounted, QALYs	Hypothetical Comparator Costs - discounted, £M	Hypothetical Comparator Effects - discounted, QALYs
1	45	45	$(45 \cdot 10^6)/20{,}000 =$ 2,270	45	$(45 \cdot 10^6)/20{,}000 =$ 2,270
2	202	202	$(202 \cdot 10^6)/20{,}000 =$ 10,100	$\delta_C^{\ 1} \cdot 202 =$ 195	$\delta_E^{\ 1} \cdot (202 \cdot 10^6)/20{,}000 =$ 9,950
3	256	256	$(256 \cdot 10^6)/20{,}000 =$ 12,780	$\delta_C^{\ 2} \cdot 256 =$ 239	$\delta_E^{\ 1} \cdot (256 \cdot 10^6)/20{,}000 =$ 12,410
4	99	99	$(99 \cdot 10^6)/20{,}000 =$ 4,950	$\delta_C^{\ 3} \cdot 99 =$ 89	$\delta_E^{\ 1} \cdot (99 \cdot 10^6)/20{,}000 =$ 4,730
5	99	99	$(99 \cdot 10^6)/20{,}000 =$ 4,950	$\delta_C^{\ 4} \cdot 99 =$ 86	$\delta_E^{\ 1} \cdot (99 \cdot 10^6)/20{,}000 =$ 4,660
…	…	…	…	…	…
100	99	99	$(99 \cdot 10^6)/20{,}000 =$ 4,950	$\delta_C^{\ 99} \cdot 99 =$ 3	$\delta_E^{\ 1} \cdot (99 \cdot 10^6)/20{,}000 =$ 1,130
101	99	99	$(99 \cdot 10^6)/20{,}000 =$ 4,950	$\delta_C^{\ 100} \cdot 99 =$ 3	$\delta_E^{\ 1} \cdot (99 \cdot 10^6)/20{,}000 =$ 1,120
Total				3,030	270,522
CE ratio (3 s.f.)					11,200

$\delta = (1+r_C)^{-1}$ = discount factor for costs, where r_C is the discount rate for costs, $\delta_E = (1+r_E)^{-1}$ = discount factor for effects, where r_E is the discount rate for effects.
QALY: Quality adjusted life year

In the case of the algebraic form of the adjustment factor, the total number of cohorts in the model is n (1 current cohort and n-1 future cohorts). An expenditure weight s_i, denotes the relative amount of spending on the intervention per cohort. For example, if the first cohort is twice as large as all subsequent cohorts and requires twice as many vaccinations, then $s_1 = 2$, while $s_{2...n} = 1$. In the case of an intervention taking longer than one year to complete per cohort, the expenditure weight is given by the present value of all expenditure on the intervention during the cohort's treatment discounted to the year it starts the intervention. The discount factors for costs and effects are given by δ_C and δ_E respectively. The adjustment factor, α, is then

$$\alpha = \frac{\sum_{i=1}^{n} s_i \cdot \delta_C^{i-1}}{\sum_{i=1}^{n} s_i \cdot \delta_E^{i-1}}. \tag{1}$$

If the distribution of expenditure is constant over all periods, the s_i expenditure weights will be uniform. The sum of the discount factors over n cohorts can then be written as the sum of a geometric series for both the numerator and denominator

$$\alpha = \frac{\sum_{i=1}^{n} s \cdot \delta_C^{i-1}}{\sum_{i=1}^{n} s \cdot \delta_E^{i-1}} = \frac{s(1-\delta_C{}^n)/(1-\delta_C)}{s(1-\delta_E{}^n)/(1-\delta_E)}, \tag{2}$$

so the adjustment factor is then given by

$$\alpha = \frac{(1-\delta_C{}^n)(1-\delta_E)}{(1-\delta_E{}^n)(1-\delta_C)}. \tag{3}$$

DISCUSSION

The results show how to adjust the current threshold for the number of future modelled. In our example the adjusted threshold is £11,200/QALY. Since the ICER of vaccination is still below this adjusted threshold we conclude vaccination is cost-effective. Alternatively, we can adjust the intervention's ICER. That the adjusted ICER of £13,300/QALY is below the current threshold reflects the conclusion that vaccination is cost-effective.

The interpretation of our results is straightforward. Our example shows the adjusted threshold for an intervention modelled with multiple future cohorts to be lower than the current threshold. This may seem counter-intuitive, however, the rationale is that a lower, more restrictive threshold must be imposed when a CEA contains multiple future cohorts to compensate for the more favourable cost-effectiveness estimate due to the inter-cohort effect.

The adjusted ICER of vaccination in our example is greater than the unadjusted ICER, but less that the ICER under equal discounting. The adjustment factor removes the inter-cohort effect of differential discounting. However, the benefit of the intra-cohort effect remains; evidenced by the lower adjusted ICER relative to the equal discounting ICER.

The adjustment factor is independent of the threshold. Consequently, it can be applied where there is no explicit threshold, but where there is a rate of threshold growth implied by differential discount rates, such as in the Netherlands. Similarly, it can be applied where the threshold is a range rather than a point value, as in the UK.

The inclusion of catch-up vaccination in our example means the expenditure weights were uneven between cohorts. Had no catch-up been assumed, the expenditure weights would be even and the unweighted adjustment factor given by equation 3 could be used. The unweighted adjustment factor is 0.54, just marginally different from the weighted adjustment factor of 0.56 from the example. Indeed, the unweighted and weighted adjustment factors will be approximately equal when the

differences in expenditure between cohorts are small. Consequently, the unweighted adjustment factor can be used to approximate the adjusted ICERs from published studies where the actual expenditure distribution is unknown but assumed close to uniform.

In the case of interventions without shared effects, the adjusted ICER from our solution can be shown to be equivalent to two-step discounting, where the second step discount rate equals the effects discount rate (see Appendix I). This result is useful, as previously it was unclear what discount rates to employ in two-step discounting, but the hypothetical comparator provides an intuitive rationale for applying discount rate on effects in the second step. In cases without indirect effects it may be more convenient to apply two-step discounting than our adjustment.

Our analysis assumes the hypothetical comparator should be matched to the present value of intervention costs per cohort in the year the cohort starts the intervention. This is consistent with comparisons of interventions modelled in single cohorts to the threshold, as currently commonly practiced in CEA. It could be argued that to truly account for the opportunity cost of the intervention of interest we should match the hypothetical comparator to costs in each period in which they are incurred rather than just to their present value at the start year. While this would not affect our example, as the vaccination costs for each cohort are only incurred in one year for each cohort, it would change the adjustment factor in an intervention with costs lasting longer than one year per cohort.

An explicit premise of our analysis is that interventions should only be accepted as cost-effective with respect to the current threshold and not because of anticipated threshold growth in periods when future cohorts start the intervention. According to this view, if the intervention is not yet cost-effective its implementation should be postponed until the threshold has increased sufficiently.

This premise, however, may be disputed by those who consider the average cost-effectiveness over multiple periods the appropriate basis on which to judge cost-effectiveness. Hoyle and Anderson recommend interventions be modelled over all anticipated recipient cohorts, both current and future [24]. Their recommendation is

in part motivated by the realisation that differential discounting leads to cost-effectiveness progressively improving between cohorts. Our perspective differs to that of Hoyle and Anderson. We contend that in most cases it is unnecessary to make a policy choice to reimburse an intervention for its entire potential implementation lifetime at one point in time; rather, it is possible to review reimbursement as required.

In summary, the proposed adjustment provides a simple, pragmatic and transparent way to adjust estimates from CEA models using multiple future cohorts under differential discounting. The logic of our adjustment is simple and easy to follow; it places no restrictions on the number of future cohorts modelled; and, is both more intuitive and more general than two-step discounting. While the adjustment factor can be applied to either the threshold or the intervention's ICER, adjusting the ICER is more convenient, as it permits direct comparisons between interventions and to the current cost-effectiveness threshold.

Limitations

A limitation of the analysis presented here is that it does not consider uncertainty. The cost-effectiveness threshold is treated deterministically in this analysis in that the cost-effectiveness of the comparator intervention is not assumed to be a random variable. Whereas the cost-effectiveness threshold interpreted as the cost-effectiveness of the marginal intervention not adopted at the margin is not known with certainty, as the cost-effectiveness of the marginal intervention itself is uncertain [34]. Indeed, this uncertainty is reflected in part by NICE's use of a threshold range rather between than a point value. However, it would be possible to apply the hypothetical comparator twice, once with the lower bound and once with the upper. This would allow the calculation of threshold range for the adjusted threshold with which to assess the intervention of interest.

CONCLUSION

Differential discounting prevents direct comparisons between CEAs with different numbers of cohorts and to the current cost-effectiveness threshold. The proposed adjustment based on the hypothetical comparator allows analysts to retain the intra-cohort effect of differential discounting, while removing the inter-cohort effect that confuses comparisons. Consequently, using the proposed adjustment, CEAs can still account for an increasing value of health over time, but avoid confusion regarding the practical implementation of differential discounting in models with multiple cohorts.

APPENDIX: Equivalence of the Hypothetical Comparator to Two-Step Discounting

The adjusted ICER derived from the hypothetical comparator method is shown here to be equivalent to the cost-effectiveness ratio derived from a two-step discounting procedure. In two-step discounting's first step, costs and effects are discounted to the intervention start year for each cohort, using the discount factors for costs and health effects respectively. In the second step, the present values of costs and effects at the intervention start years are discounted back to the common discount year using the discount factor for health effects.

The following example demonstrates the equivalence of two-step discounting to the hypothetical comparator method in a simple example with no indirect effects. First we show the adjusted ICER from the hypothetical comparator method. We then apply two-step discounting to the same costs and effects and show that the resulting cost-effectiveness ratio is the same as the adjusted ICER.

An intervention taking two periods for a given cohort to complete is implemented in two consecutive cohorts. The costs and effects for each cohort in each period are shown in Table A1 below. The C and E terms represent costs and effects. The subscripts indicate whether they relate to the first or second year of implementation for each given cohort.

Table A1. Costs and effects of an intervention implemented in two cohorts over three periods

	Period		
Cohort	1	2	3
1	C_1	C_2	
	E_1	E_2	
2		C_1	C_2
		E_1	E_2

The intervention's discounted costs and effects of the intervention can be written as

$$C_1 + \delta_C C_2 + \delta_C C_1 + \delta_C{}^2 C_2, \text{ and} \tag{1}$$

$$E_1 + \delta_E E_2 + \delta_E E_1 + \delta_E{}^2 E_2. \tag{2}$$

respectively, where δ_C and δ_E are the discount factors for costs and effects respectively. This yields the following unadjusted cost-effectiveness ratio

$$\frac{(1+\delta_C)(C_1 + \delta_C C_2)}{(1+\delta_E)(E_1 + \delta_E E_2)}. \tag{3}$$

For each cohort, the present value of costs in the period they start the intervention is given by

$$C_1 + \delta_C C_2. \tag{4}$$

For the hypothetical intervention we assume costs occurring in periods 1 and 2 corresponding to cohorts 1 and 2. These costs are equal to the present value of the intervention costs in each intervention start year given as

$$C_1 + \delta_C C_2. \tag{5}$$

The total discounted cost of the hypothetical comparator equals the present value in the discount year of the costs of the hypothetical intervention matched to two cohorts starting the intervention one period apart, which can be written as

$$(1+\delta_C)(C_1 + \delta_C C_2). \tag{6}$$

The health effects for each cohort of the hypothetical intervention are obtained dividing the costs of the hypothetical intervention by the cost-effectiveness threshold

$$(C_1 + \delta_C C_2)/T. \tag{7}$$

The hypothetical comparator's total discounted health effects in the discount year can be written as

$$(1+\delta_E)(C_1+\delta_C C_2)(1/T). \tag{8}$$

The hypothetical comparator's cost-effectiveness ratio, which serves as the adjusted cost-effectiveness threshold, can be then be written as

$$T\frac{1+\delta_C}{1+\delta_E}\frac{C_1+\delta_C C_2}{C_1+\delta_C C_2}. \tag{9}$$

Dividing the adjusted threshold by the current threshold yields the adjustment factor

$$\frac{1+\delta_C}{1+\delta_E}. \tag{10}$$

Finally, dividing the unadjusted cost-effectiveness ratio (equation 3) by the adjustment factor yields the adjusted cost-effectiveness ratio

$$\frac{C_1+\delta_C C_2}{E_1+\delta_E E_2}. \tag{11}$$

Now we apply two-step discounting to the intervention as described above. The costs and effects of the intervention in the discount year are now respectively written as

$$C_1+\delta_C C_2+\delta_E(C_1+\delta_C C_2) \text{ and} \tag{12}$$

$$E_1+\delta_E E_2+\delta_E(E_1+\delta_E E_2). \tag{13}$$

Dividing equations 12 by 13 and simplifying gives the cost-effectiveness ratio

$$\frac{C_1+\delta_C C_2}{E_1+\delta_E E_2}, \tag{14}$$

which equals the adjusted cost-effectiveness ratio given by equation 11. This simple two period proof can be extended to multiple periods and the same equivalence between the two methods will be observed.

REFERENCES

1. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics. 2001;10(7):587-99.

2. Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal. 2005;331(7514):446-8.

3. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press; 2006.

4. Gravelle H, Brouwer W, Niessen L, Postma M, Rutten F. Discounting in economic evaluations: stepping forward towards optimal decision rules. Health Economics. 2007;16(3):307-17.

5. Westra TA, Parouty M, Brouwer WB, Beutels PH, Rogoza RM, Rozenbaum MH, et al. On Discounting of Health Gains from Human Papillomavirus Vaccination: Effects of Different Approaches. Value Health. 2012;15(3):562-7.

6. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health care technologies. Health Econ. 2011;20(1):2-15.

7. Cleemput I, van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Brussels: Health Care Knowledge Centre (KCE), 2008.

8. AOTM. Guidelines for conducting Health Technology Assessment (HTA). Warsaw: Agencja Oceny Technologii Medycznych, 2009.

9. Oostenbrink JB, Bouwmans CAM, Koopmanschap MA, Rutten FFH. Guidelines for Pharmacoeconomic Research, Updated Version. Diemen: College voor Zorgverzekeringen, 2004.

10. NICE. Discounting of Health Benefits in Special Circumstances. National Institute for Health and Clinical Excellence, 2011.

11. Koleva D, De Compadri P, Padula A, Garattini L. Economic evaluation of human papilloma virus vaccination in the European Union: a critical review. Internal and Emergency Medicine. 2011;6(2):163-74.

12. Demarteau N, Detournay B, Tehard B, El Hasnaoui A, Standaert B. A generally applicable cost-effectiveness model for the evaluation of vaccines against cervical cancer. Int J Public Health. 2011;56(2):153-62.

13. Mennini FS, Giorgi Rossi P, Palazzo F, Largeron N. Health and economic impact associated with a quadrivalent HPV vaccine in Italy. Gynecol Oncol. 2009;112(2):370-6.

14. Lazaro A. Theoretical arguments for the discounting of health consequences: Where do we go from here? PharmacoEconomics. 2002;20(14):943-61.

15. Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. Management Science. 1983;29(3):300-6.

16. van Hout B. Discounting costs and effects: a reconsideration. Health Economics. 1998;7(7):581-94.

17. O'Mahony JF, de Kok IMCM, van Rosmalen J, Habbema JDF, Brouwer W, van Ballegooijen M. Practical implications of differential discounting in cost-effectiveness analyses with varying numbers of cohorts. Value in Health. 2011;14(4):438-42.

18. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of Human Papillomavirus vaccination in the Netherlands. Journal of the National Cancer Institute. 2009;101(15):1083-92.

19. de Kok IMCM, van Ballegooijen M, Habbema JDF. Response: Re: Cost-Effectiveness Analysis of Human Papillomavirus Vaccination in the Netherlands. Journal of the National Cancer Institute. 2010;102(5):358.

20. Boot HJ, Wallenburg I, de Melker HE, Mangen M-JM, Gerritsen AAM, van der Maas NA, et al. Assessing the introduction of universal human papillomavirus vaccination for preadolescent girls in The Netherlands. Vaccine. 2007;25(33):6245-56.

21. Coupé VMH, van Ginkel J, de Melker HE, Snijders PJF, Meijer CJLM, Berkhof J. HPV16/18 vaccination to prevent cervical cancer in The Netherlands: Model-based cost-effectiveness. International Journal of Cancer. 2009;124(4):970-8.

22. Rozenbaum MH, Boersma C. Response: Re: Cost-Effectiveness Analysis of Human Papillomavirus Vaccination in the Netherlands. Journal of the National Cancer Institute. 2010;102(5):358-9.

23. Postma, Maarten J. Cost-effectiveness analysis of Human Papillomavirus (HPV) Vaccination in the Netherlands: Recent publication reinforces favorable cost-effectiveness despite misleading conclusion. Vaccine. 2010;28(4):873-4.

24. Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Medical Decision Making. 2010;30(4):426-37.

25. Brisson M, Edmunds WJ. Economic evaluation of vaccination programs: The impact of herd-immunity. Health Econ. 2003;23(1):76-82.

26. Jit M, Brisson M. Modelling the Epidemiology of Infectious Diseases for Decision Analysis: A Primer. PharmacoEconomics. 2011;29(5):371-86.

27. Salminen E, Kiel K, Ibbott G, Joiner M, Rosenblatt E, Zubizarreta E, et al. International Conference on Advances in Radiation Oncology (ICARO): Outcomes of an IAEA Meeting. Radiat Oncol. 2011;6(11):1-9.

28. Gyrd-Hansen D, Søgaard J. Discounting life-years: whither time preference? Health Economics. 1998;7(2):121-7.

29. Lipscomb J. Time Preference for Health in Cost-Effectiveness Analysis. Medical Care. 1989;27(3):S233-S53.

30. Lipscomb J, Weinstein MC, Torrence GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996. p. 214-46.

31. Jit M, Chapman R, Hughes R, Choi YH. Comparing bivalent and quadrivalent human papillomavirus vaccines: economic evaluation based on transmission model. Br Med J. 2011;343:1-15.

32. Jit M, Choi YH, Edmunds WJ. Economic evaluation of human papillomavirus vaccination in the United Kingdom. Br Med J. 2008;337:331-45.

33. McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold: What it is and What that Means. PharmacoEconomics. 2008;26(9):733-44.

34. Sendi P, Gafni A, Birch S. Opportunity costs and uncertainty in the economic evaluation of health care interventions. Health Econ. 2002;11(1):23-31.

Paper 3.

Describing the Implementation Time Horizon in Cost-Effectiveness Analysis and Assessing its Impact Using the Net Benefit Framework

James F. O'Mahony, MA[1,2],
Joost van Rosmalen, PhD[2],
Ann G. Zauber, PhD[3],
Marjolein van Ballegooijen, MD, PhD[2].

1. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland;
2. Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands;
3. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Centre, New York, USA.

ABSTRACT

Introduction. The role of time horizons for costs and effects in cost-effectiveness analysis (CEA) is well understood. However, one aspect of time horizons has not been described previously. It is the length of time over which the implementation of an intervention is modelled; described here as the implementation horizon. When using models with multiple birth cohorts to simulate interventions such as cancer screening programmes that take many years for a given cohort to complete, finite implementation horizons can result in younger cohorts stopping the intervention before the screening stop age. A concern is that this premature cessation compromises the external validity of such models. This study describes the implementation horizon and assesses the impact of a 30-year horizon used in a recently published CEA of colorectal screening strategies in the Netherlands. It also considers shorter horizons of 20 and 10 years.

Methods. The colorectal screening model used in the recent Dutch CEA is run with and without the 30-year implementation horizon. The differences in the optimal strategies between the two versions of the model are compared in terms of net health benefit. The analysis is repeated for the 20 and 10-year horizons.

Results. The optimal screening strategies differ very little with and without the horizon, which is reflected in only minor differences in net health benefit between the models. This indicates that a 30-year horizon does not meaningfully distort the optimal choices of screening strategy. Only when a 10-year horizon is imposed is there a significant distortion to the optimal choices.

Conclusion. Despite censoring a large proportion of the screens in the model, a 30-year implementation horizon does not result in a strong systematic bias among the alternative screening schedules in this example. However, the impact of a shorter horizon demonstrates that implementation horizons can compromise external validity.

INTRODUCTION

This study addresses the length of time interventions are implemented in cost-effectiveness models. It considers the simulation of interventions that take many years for a given cohort to complete in models with multiple birth cohorts. In particular, this study investigates the practice of assuming implementation ceases at a given point in time for all cohorts, irrespective of whether the intervention has been concluded by all cohorts. This is an aspect of model time horizons that has not been previously addressed by the literature. This paper illustrates the issue using a recently published model-based CEA of colorectal screening in the Dutch population over a period of 30 years.

Background to the Implementation Time Horizon

Models are often used in CEAs of healthcare interventions. They allow the synthesis of data from multiple sources, the extrapolation of outcomes beyond trial follow-up periods and the comparison of multiple intervention alternatives that have not or cannot be compared in trials [1,2]. The variety of modelling approaches in CEA has been described according to various taxonomies [3-7]. One categorical distinction is between cohort and population models. While CEAs often simulate a single cohort of patients, some feature multiple cohorts. Multiple cohort models sometimes feature incident cohorts that enter the model over time, either as they are born, develop a condition or reach a certain age. Such multiple cohort models are often described as population models [8].

The scope of time within CEA models is bounded by the time horizon. The time horizon is the point in time at which the assessment of costs and effects in the model ceases. Methods guidelines suggest the time horizon should be sufficiently long to capture all meaningful differences in costs and effects due to an intervention [9]. This often implies the assessment of costs and effects until death [10].

The appropriate time horizon requires particular consideration in the case of multi-cohort models. Karnon et al. discuss how the entry of incident cohorts into a model requires the time horizon to be longer than if no incident cohorts were included if all costs and effects are to be captured [11]. Alternatively, cross-sectional

interpretations of population models with time horizons that censor the assessment of costs and effects for some or all cohorts have been described in the literature, whereby costs and effects are assessed over a finite time span irrespective of whether the cohorts remain alive or not [4,12,13].

The cross-sectional approach of assessing all costs and effects within a horizon does not accord with the conventional longitudinal approach applied in CEA of assessing the lifetime (or other relevant span) costs and effects for each cohort. Consequently, although the cross-sectional approach is useful for budget impact analysis, it is not appropriate for cost-effectiveness analysis, as it neglects to account for costs and effects that occur after the time horizon [11,12,14,15].

Despite the understanding of the appropriate time horizons for costs and effects in the literature, there is another form of time horizon sometimes used in CEA that has not previously been described or assessed. It applies to the specific case of multi-cohort models of interventions that take longer than one year for a given cohort to complete. Cancer screening programmes are examples of such interventions, as a complete screening programme may comprise multiple screening moments and take decades for a single cohort to complete. When these long-duration interventions are modelled in multiple cohorts, the result is multiple overlapping recipient cohorts, as each cohort is at a different point of their screening schedule at any given point in time.

Some multi-cohort models impose a finite horizon on the implementation of the intervention, meaning the intervention ceases for all cohorts after a period of 20 or 30 years, but the assessment of health effects and cost savings is continued until the death of all cohorts. Imposing an implementation horizon on multi-cohort models of long-duration interventions can result in some simulated cohorts concluding the intervention before they reach the screening stop age. When implementation horizon does constrain cohorts from completing the intervention in this way we describe it as a binding implementation horizon and the model as a constrained model. Conversely, when the implementation horizon is sufficiently long as not to constrain any cohorts from completing the intervention it is described as a non-binding implementation horizon and the model is described as an unconstrained model.

In the case of an intervention that takes longer than one year for a cohort to complete, when all cohorts starting the intervention within the implementation horizon are included in the model, at least one cohort will have its implementation constrained by the horizon. The longer the intervention takes to complete for a single cohort, the greater the number of cohorts will have their intervention curtailed by a given horizon.

Figure 1A and B shows a binding implementation horizon and illustrates how it affects two different screening strategies. Panels 1A and B illustrate screening between ages 60-69 and 45-80 respectively, which are the shortest and longest screening strategies considered in the example used in this paper. Each horizontal line represents the lifetime of each cohort, entering the model at birth and leaving at death. The oldest cohort in this model is born in 1925 while the youngest is born in 1990. Death is represented at a single age in the diagram for convenience, but varies between individuals. The first and second vertical lines (from left to right) represent the introduction of screening and the implementation horizon respectively. In both cases screening is introduced in 2005 and ceases for all cohorts in 2035 due to the binding implementation horizon.

Some of the older cohorts are already within the screening age range when screening is introduced. Following Hoyle and Anderson's terminology we describe these cohorts as the prevalent cohorts [16]. One cohort will be at the screening start age when screening is introduced, described as the current incident cohort. Younger cohorts yet to enter the screening age range are described as future incident cohorts.

The dark grey section of each line represents the period over which each cohort is screened. Periods of screening censored by the binding implementation horizon are shown with the lighter grey sections. Different screening strategies will feature more or less prevalent cohorts, depending on the strategy's screening age range. Similarly, the portion of the strategy censored by the implementation horizon will also vary. Figure 1 shows both that the implementation horizon censors a very large proportion of screening in the two strategies and that the absolute amount of screening that is censored is much greater in the 45-80 strategy.

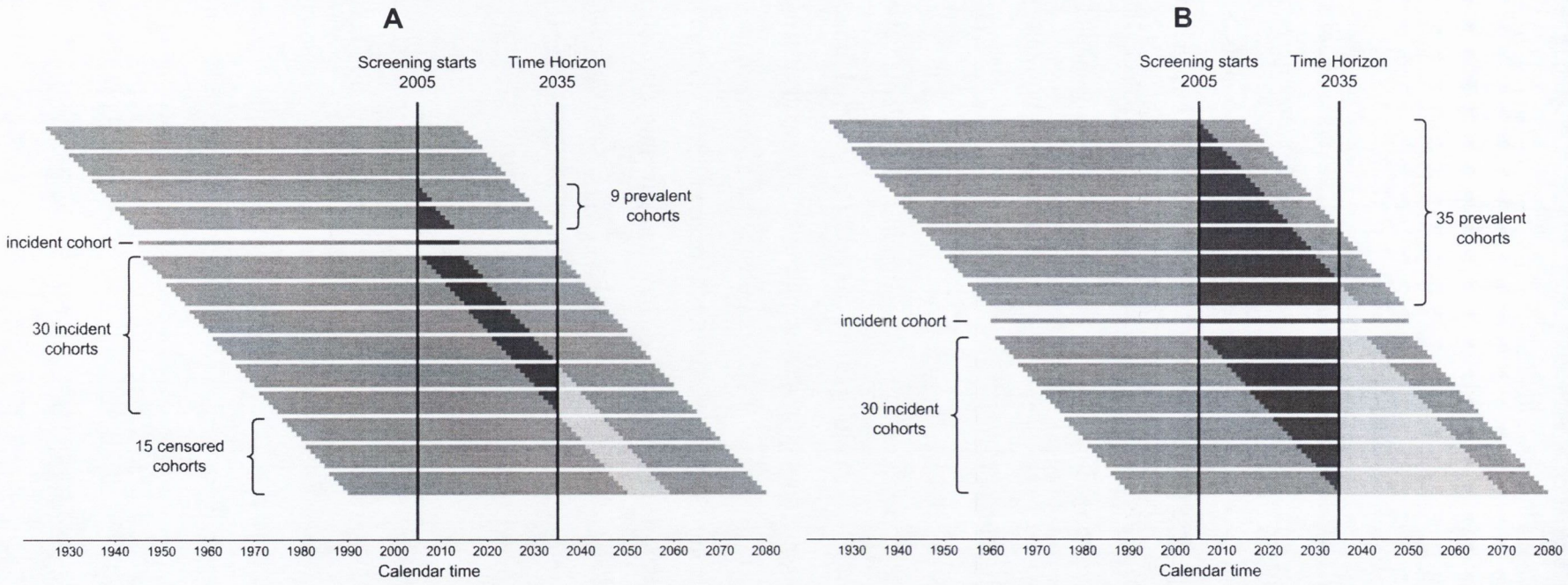
Note that the 15 youngest cohorts in the 60-69 strategy are completely censored by the implementation horizon. Conversely, while none of the cohorts under the 45-80 strategy are completely censored by the horizon, none actually complete the full screening programme either, as the implementation horizon is shorter than the complete screening strategy itself.

Concerns of Bias

We do not expect all screening moments within a given strategy to be equally cost-effective. The cost-effectiveness of screening will vary within a given cohort with disease incidence and remaining life expectancy, both of which vary with age. Similarly, screening cost-effectiveness can vary within a population depending on its previous screening history, due to differences in the prevalence of undetected disease.

The extensive censoring of screens by a binding implementation horizon and the expectation of varying cost-effectiveness between screens prompts concerns of a potential for bias. The varying cost-effectiveness between screens means an implementation horizon may influence cost-effectiveness estimates, as the cost-effectiveness of screens censored by the horizon may differ from that of screens occurring within the horizon. If it is assumed that in reality all cohorts will complete screening rather than stopping prematurely, then analysis using a constrained model will be unrepresentative of reality. As such, if the model is considered unrepresentative of a likely reality, it is considered to have poor external validity.

Figure 1: Multi-cohort model featuring 66 consecutive annual birth cohorts born since 1925 illustrating screening between the ages of (A) 60-69 and (B) 45-80 introduced in 2005 with a binding implementation horizon of 30 years: each birth cohort represented by a separate grey bar starting in the year of the cohorts birth, with periods of screening provision indicated by the dark section of each bar, with periods of screening provision censored by the implementation horizon indicated with light grey sections of each bar



Examples from the Literature

The literature contains examples of CEA models of cancer screening that simulate multiple cohorts and feature binding implementation horizons while assessing all costs and effects until death. A number of these examples feature the MISCAN microsimulation model of cancer prevention developed at the Erasmus Medical Centre in Rotterdam, the Netherlands. These examples include Wilschut et al., Loeve et al. and Ramsey et al., which are all analyses of colorectal cancer screening that impose 30-year implementation horizons [17-19]. De Gelder et al. uses a multi-cohort model with a 20-year implementation horizon in its assessment of breast cancer screening [20]. Boer et al., de Koning et al., Groenewoud et al. and van der Maas et al. all employ 27-year implementation horizons in their assessments of breast cancer screening [21-24]. Similarly, Koopmanschap et al. and van den Akker-van Marle et al. use 27-year implementation horizons when assessing cervical screening [25,26].

There are also examples of CEAs with binding implementation horizons from other than the MISCAN modelling group. Gyrd-Hansen et al. imposes a 36-year implementation horizon when comparing multiple cervical screening strategies [27]. Szeto and Devlin uses a 30-year implementation horizon in an assessment of mammography [28]. Brown uses a costs and effects horizon of 30 years, but uses an implementation horizon of 20 years when assessing mammography screening [29]. Lejenue et al. and Lejeune et al. both use a 20-year implementation horizon when assessing colorectal screening [30,31]. Stout et al. assesses mammography screening using an implementation horizon of 10 years [32].

Analysis Overview

The study assesses the impact of binding implementation horizons on optimal policy choices. This assessment uses a model employed in a recently published CEA of colorectal screening in the Dutch population by Wilschut et al. [17]. Wilschut et al. compared 48 different combinations of screening start ages, stop ages and screening intervals and 4 quantitative cut-off levels of a faecal immunochemical test (FIT) for population screening in the Netherlands. The model simulated a population comprising 66 birth-year cohorts to match the potential screen recipients in the Dutch population over the next 30 years. All cohorts ceased

screening at a 30-year implementation horizon, irrespective of whether they had reached the screening stop age. The assessment of all costs and health effects continued for all cohorts until death.

We found the screening strategies identified as lying on the efficient frontier in the cost-effectiveness plane by the constrained model with the 30-year implementation horizon. We then found the strategies identified as efficient by the unconstrained model without the 30-year horizon, which allows all cohorts to continue screening until screening the stop age. The efficient strategies from the constrained and unconstrained models were then compared using the net benefit framework [33]. The net benefit framework allows us to quantify the loss in efficiency caused by constraining the model with the implementation horizon.

METHODS

We repeated the analysis presented in Wilschut et al. using the same CEA model. This validated model has previously been described in Wilschut et al. and other publications including the standardised MISCAN model description [34-36]. We assessed the same 48 screening strategies, comprising combinations of screening intervals of 1, 1.5, 2, 3 years with start and stop ages of 45, 50, 55, 60 and 70, 75 and 80 respectively. Our analysis only considered the FIT cut-off level of 50 nanogrammes of haemoglobin/millitre as Wilschut et al. found this to dominate all others.

We simulated the life histories of approximately 75 million people born between 1925 and 1990. We increased the size of the simulated population used in the original analysis in order to attenuate simulation error. The size of each birth year cohort is proportionate to that in the original CEA. The results have been rescaled to report the estimated costs and effects for a population of 100,000 people.

We appraised the difference between the efficient strategies identified by constrained and unconstrained models over a range of cost-effectiveness threshold values in terms of net health benefit (NHB). Equation 1. gives the expression for net

health benefit, where E is the estimated health effect of the intervention, C is the estimated cost and λ is the cost-effectiveness threshold.

$$NHB = E - \frac{C}{\lambda} \quad (1)$$

The strategy with the greatest NHB at a given level of the threshold is considered optimal at that threshold. The set of optimal strategies over a range of threshold values make up the efficient frontier.

We identified the optimal strategies according to the constrained model. We estimated the NHB of these strategies when simulated over a range of threshold values in the unconstrained model. We then estimated the NHB of the optimal policy choices from the unconstrained model in the unconstrained model itself. We express the NHB of the strategies identified as optimal by the constrained model as a ratio of the NHB of the optimal strategies from the unconstrained model over a range of threshold values. This ratio indicates how much of the potential optimal net benefit from the unconstrained model is achieved by the choices from the constrained model. We also repeated this analysis for shorter implementation horizons of 20 and 10 years.

The underlying assumption of the comparison of net benefit described above is that the unconstrained model is representative of reality while the constrained model is not, as in reality screening will not cease at the implementation horizon. Since the strategies identified as optimal by the constrained analysis are from a model not assumed to be representative of reality, they may not necessarily be optimal. Estimating the difference in net benefit between the optimal choices from the constrained model to the optimal choices from the unconstrained model in the unconstrained model itself allows us to quantify any implied inefficiency of failing to choose the optimal strategies as a result of using the constrained model.

RESULTS

Table 1 below reports the incremental cost-effectiveness ratios (ICERs) of the efficient strategies from the model run with and without the implementation horizon. Not all strategies are common to both frontiers, but for those strategies that are common the ICERs are consistently somewhat higher in the unconstrained model. The efficient strategies and ICERs in Table 1 do not exactly match those reported by Wilschut et al. because of the difference in stochastic simulation error between the models due to the differences in the size of the simulated populations. A full table of all cost and effects estimates for all 48 strategies in the constrained and unconstrained model is given in Appendix I.

Table 1: Strategy description and estimated incremental cost-effectiveness with and without a binding 30 year implementation horizon

Strategy Age range : Interval : No. of Screens	30-year Horizon ICER, €/LYG	No Horizon ICER, €/LYG
60 – 69 : 3 : 4	1,640	1,640
60 – 70 : 2 : 6	2,520	2,620
55 – 73 : 3 : 7	2,690	2,910
55 – 75 : 2 : 11	3,080	3,330
55 – 79 : 2 : 13	ED	5,610
55 – 79 : 1.5 : 17	5,320	6,070
50 – 80 : 2 : 16	5,370	ED
50 – 80 : 1.5 : 21	6,150	7,500
50 – 80 : 1 : 31	10,610	13,860
45 – 80 : 1 : 36	16,760	56,250

ED: strategy subject to extended dominance

Figure 2. shows the cost-effectiveness plane featuring the efficient interventions from the constrained and unconstrained models. The estimates from the constrained and unconstrained models are shown with the white dots and black diamonds respectively, each joined to show the efficient frontier for both models.

Figure 2: Cost-effectiveness plane showing the efficient strategies in the constrained and unconstrained models and cost-effectiveness threshold lines indicating the net benefit of alternative screening strategies for a population of 100,000 people: A representing the optimal strategy from the constrained model at the illustrated threshold; B representing the same strategy in the unconstrained model; C representing the optimal strategy at the same threshold in the unconstrained model

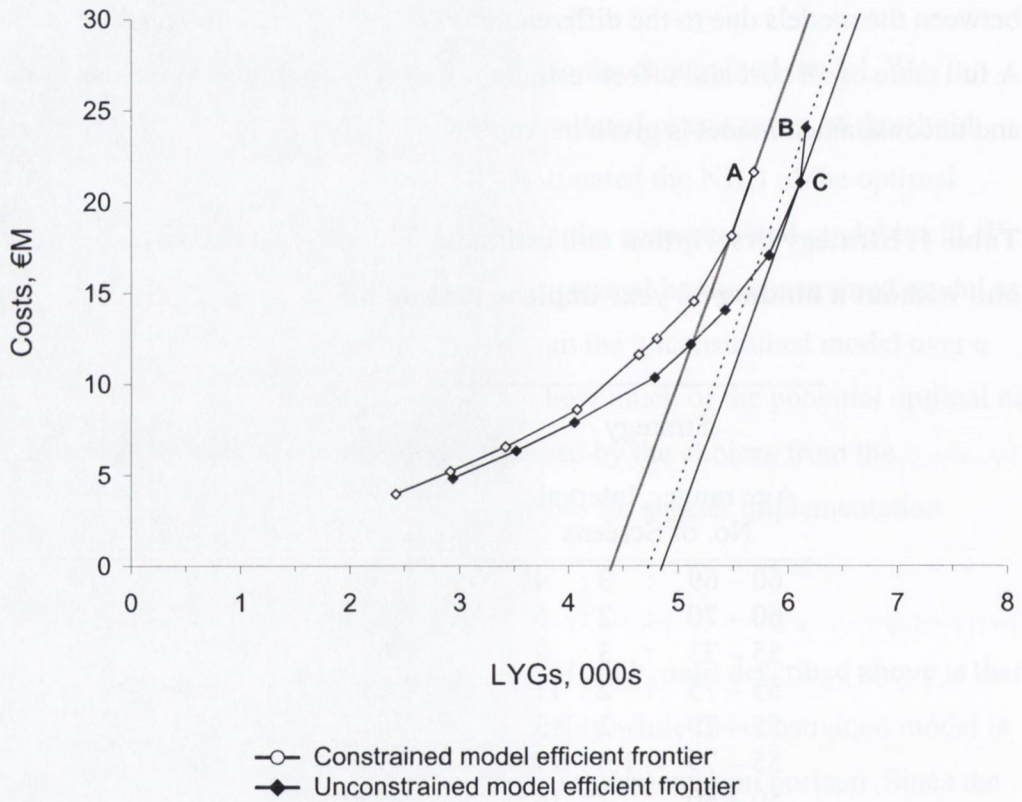


Figure 2 features three lines parallel to a cost-effectiveness threshold of €16,800/LYG to illustrate the estimation of NHB. At that threshold, the strategy marked A has the highest net benefit in the constrained model. That strategy modelled in the unconstrained model is shown with point B. The optimal strategy in the unconstrained model at the same threshold is marked with point C. The NHB of each strategy at that threshold is given by the intersection of each threshold line with the horizontal axis. The difference in NHB of the optimal choices from the two models at this threshold when simulated in the unconstrained model is shown by the difference between the dotted and solid black threshold lines at the intersection with the horizontal axis.

Figure 3: Net monetary benefit of efficient screening strategies when modelled with binding implementation horizon of 30 and 20 years relative to net benefit of efficient strategies when modelled without an implementation horizon over a range of cost-effectiveness threshold values

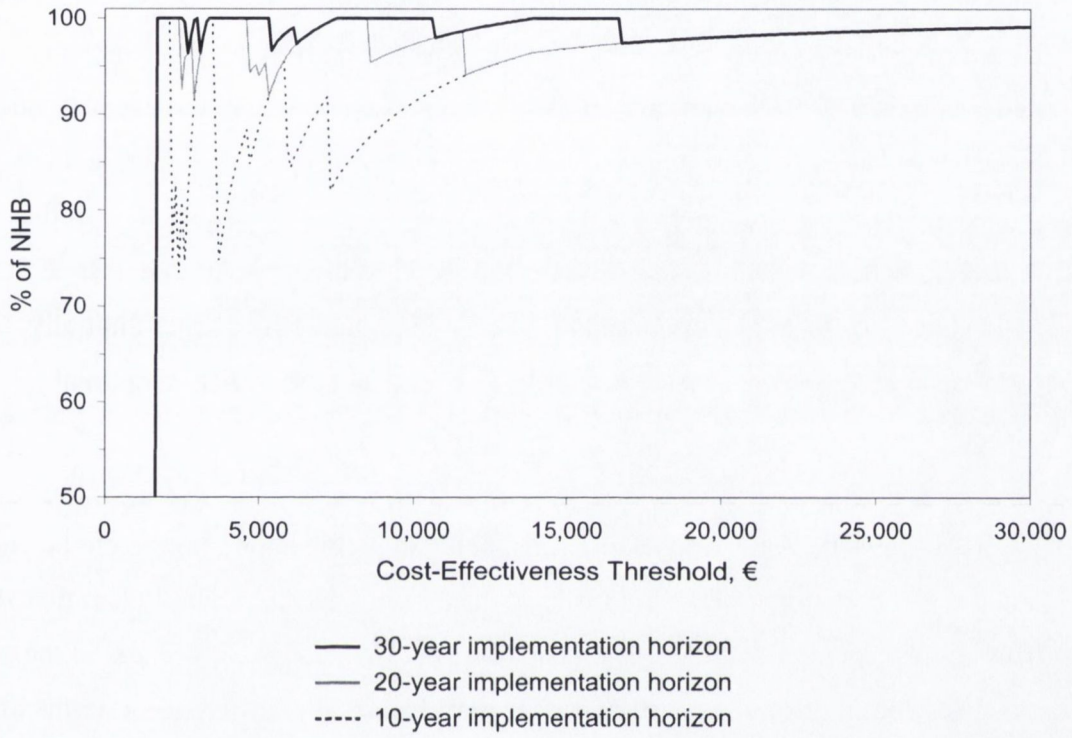


Figure 3 shows the NHB of the optimal strategies according to the constrained model when assessed in the unconstrained model as a percentage of the NHB of the optimal strategies from the unconstrained model over a range of threshold values. The black line corresponds with the 30-year horizon used in the published CEA. The grey and dotted lines shows the result for the same analysis repeated with 20 and 10-year implementation horizons respectively. The figure shows that the optimal strategies from the constrained model achieve only marginally less NHB than the optimal schedules from the unconstrained model over a wide range of threshold values: generally, the choices of the constrained model yield in excess of 96% of the NHB of the optimal choices of the unconstrained model. At some threshold values the optimal strategies selected by the two models are the same, resulting in a 100% net benefit ratio. The pattern for the 20 and 10-year horizons is broadly similar, but the losses of net benefit resulting from the implementation horizon are larger.

DISCUSSION

Interpreting the Results

Table 1 shows that the optimal screening strategies from the model with and without the implementation horizon are similar, as most strategies are common to both models. The ICERs estimated from the constrained model are consistently marginally lower in the common efficient strategies compared to the unconstrained model; consequently the constrained model will always suggest a shift to a more intensive screening strategy at a somewhat lower value of the threshold than the unconstrained model. This is reflected in the shape of Figure 3. The choices of the constrained model switch prematurely, resulting in a loss of net benefit initially, followed by convergence as the threshold increases and the choice of optimal strategies converge.

Figure 3 shows the loss in NHB of imposing the implementation horizon to be very small. Over the range of threshold values, the loss in NHB is generally less than 4%. This shows that although the implementation horizon leads to differences in the choices of optimal strategy at different threshold levels, the difference in terms of expected NHB are minimal. Although no uncertainty analysis was considered in this analysis, the differences in NHB between the optimal choices due to the 30-year implementation horizon are so small that it is likely that they would be dwarfed by error bounds around the frontier resulting from an uncertainty analysis.

Figure 3 shows that the imposition of shorter binding implementation horizons will lead to greater loss in NHB. However, the figure shows that the losses remain modest with a 20 implementation horizon and only become considerable with the shorter horizon of 10 years.

These results provide reassurance that the results presented in Wilschut et al. did not provide decision makers with misleading estimates of which strategies were optimal. Clearly, it would be mistaken to assume that implementation horizons do not matter in all situations. Indeed, the results for the 10-year horizon show that a shorter implementation horizon can lead to strategies that result in markedly less

NHB than an unconstrained analysis. Furthermore, the impact of implementation horizons may be greater in other interventions.

Why the Loss of NHB is so Modest

It is somewhat surprising that the 30-year implementation horizon made such little difference to the optimal screening choices, given that the proportion of screens censored is so large. However, the small differences between the models can be explained.

A simple explanation for the small differences between the constrained and unconstrained models would be that discounting means the present value of costs and effects of screens 30 years or more after the discount year are small relative to those of screens within the horizon. Indeed, the small difference in costs and effects between the constrained and unconstrained models is evident in Figure 2 as the efficient frontiers from both models lie very close to each other. Since the present value of screens beyond the implementation horizon have little weight, their presence or absence would make little difference. However, discounting does not account for the modest difference in net benefit between the models, as similarly small differences in NHB are also found when undiscounted outcomes are considered (results not shown).

The lack of any great difference in the optimal choices between the constrained and unconstrained models is simply because the differences in cost-effectiveness among screening strategies within the implementation horizon are broadly representative of those when the implementation horizon is relaxed. For example, the cost-effectiveness of increasing the screening frequency or raising the stop age are broadly consistent between the constrained and unconstrained models. The only notable exception where there appears to be consistent and systematic differences between the constrained and unconstrained models is in the case of changes to the screening start age. Holding the screening frequency and stop age constant while extending the screening start age below age 50 to 45 results in considerably lower cost-effectiveness ratios in the constrained model than in the unconstrained model. For example, in the case of screening every 1.5 years and stopping at age 69 the cost-effectiveness ratio of reducing the screening start age to 45 is €5,200/LYG in

the constrained model while it is €32,200/LYG in the unconstrained model. Similar differences are observed at different screening stop ages and intervals.

While large differences in cost-effectiveness ratios can be found between the constrained and unconstrained models in particular cases, the overall loss of net benefit is modest when all 48 strategies are considered, as shown by Figure 3. There are two principal reasons for this. The first is that while large differences in cost-effectiveness are found between the model in the particular case of dropping the screening start age to 45, there is only one such change which features in the efficient frontiers shown in Figure 2 (the change between the final two strategies in Table 1): all the other changes that make up the frontier are those which do not result in large differences between the models. Secondly, although the differences in cost-effectiveness ratios between the constrained and unconstrained models can be large, the differences in terms of costs and health effects are small, therefore the differences in NHB are also small.

Policy Relevance and Reflection on Model Assumptions

Decision makers might naturally request cost-effectiveness estimates for long-duration interventions over finite periods to correspond to a policy-relevant timeframe. In response analysts might oblige and apply a binding implementation horizon. Implementation horizons appear inconsistent from a theoretical perspective as they mix longitudinal and cross-sectional approaches by limiting implementation to a certain time period, while continuing assessment of costs and effects until death. However, the results of this analysis show that whether they matter in terms of policy choices is contingent: the 30-year horizon made no meaningful difference to policy choices, while a 10-year horizon would.

This analysis used the unconstrained model as the representation of what costs and effects would be realised given that it is unlikely that screening would simply cease for all cohorts at the implementation horizon. However, it is likely that screening technology will change over the next 30 years. Consequently, assuming the current screening technology remains in place is not likely to be representative of actual screening either. However, as future screening technology cannot be anticipated, it is imponderable to consider what might replace the current technology.

The use of the unconstrained model in this analysis to estimate the potential NHB carries an assumption that it is the correct model of screening to apply. It assumes that the same screening policy will be applied to all cohorts. Since a common screening policy may not be optimal for all cohorts, further gains could be realised by optimising screening for cohorts separately [37]. Consequently, the unconstrained model could itself be interpreted as an incorrect model to apply, as it does not allow for cohort-specific screening strategies. However, the purpose of this modelling exercise was to illustrate the impact of the implementation horizon. So while the unconstrained model of screening may itself have shortcomings, it is the appropriate comparator with which to illustrate the impact of a binding implementation horizon.

CONCLUSION

While time horizons for costs and effects are well understood in the literature, the implementation horizon has not been described previously. The implementation horizon's mixing of longitudinal and cross-sectional approaches is inconsistent from a theoretical perspective. This analysis shows the practical impact on optimal decision making is contingent: despite censoring a very large proportion of the total number of screening moments, a binding 30-year implementation horizon applied in a recently published CEA would not have distorted policy choices in a meaningful way, whereas a 10-year horizon would have.

APPENDIX

Table A1: Strategy description and estimated cost and effects with and without a 30 implementation horizon

Strategy				30 Year Horizon		No Horizon	
Age range	Interval	No. of Screens		Costs, €M	LYGs, 000s	Costs, €M	LYGs, 000s
45 - 70	1	26		17.52	4.82	19.10	5.19
50 - 70	1	21		14.02	4.59	15.90	5.09
55 - 70	1	16		11.00	3.92	12.47	4.69
60 - 70	1	11		7.39	3.29	9.01	3.93
45 - 75	1	31		19.38	5.24	21.30	5.66
50 - 75	1	26		15.91	5.02	18.21	5.58
55 - 75	1	21		13.35	4.29	14.95	5.21
60 - 75	1	16		9.51	3.75	11.78	4.50
45 - 80	1	36		21.59	5.49	24.03	5.94
50 - 80	1	31		18.13	5.26	21.03	5.87
55 - 80	1	26		15.26	4.81	17.91	5.51
60 - 80	1	21		12.95	4.43	14.94	4.82
45 - 69	1.5	17		12.47	4.42	14.66	4.68
50 - 69.5	1.5	14		10.87	4.11	12.39	4.61
55 - 70	1.5	11		8.45	3.68	9.97	4.27
60 - 69	1.5	7		5.53	2.66	6.71	3.38
45 - 75	1.5	21		14.53	4.81	16.99	5.28
50 - 74	1.5	17		11.83	4.01	14.28	5.09
55 - 74.5	1.5	14		10.01	4.08	11.97	4.76
60 - 75	1.5	11		7.64	3.39	9.47	4.10
45 - 79.5	1.5	24		16.92	5.02	19.08	5.54
50 - 80	1.5	21		14.81	4.77	17.05	5.45
55 - 79	1.5	17		11.59	4.31	14.04	5.04
60 - 79.5	1.5	14		9.94	3.95	11.81	4.39
45 - 69	2	13		10.59	4.08	12.37	4.33
50 - 70	2	11		8.64	3.35	10.70	4.30
55 - 69	2	8		6.57	3.09	8.00	3.77
60 - 70	2	6		5.20	2.67	6.34	3.23
45 - 75	2	16		12.38	4.52	14.54	4.93
50 - 74	2	13		10.27	3.80	12.22	4.71
55 - 75	2	11		8.58	3.76	10.34	4.42
60 - 74	2	8		6.78	3.24	8.01	3.70
45 - 79	2	18		14.29	4.62	16.27	5.15
50 - 80	2	16		12.56	4.52	14.75	5.05
55 - 79	2	13		10.15	3.79	12.16	4.66
60 - 80	2	11		9.34	3.80	10.73	4.07
45 - 69	3	9		8.22	3.37	9.61	3.74
50 - 68	3	7		6.47	3.03	7.95	3.49
55 - 70	3	6		5.53	2.81	6.86	3.36
60 - 69	3	4		3.96	2.18	4.81	2.64
45 - 75	3	11		9.35	3.61	11.52	4.29
50 - 74	3	9		8.47	3.32	10.03	4.11
55 - 73	3	7		6.39	2.78	7.90	3.67
60 - 75	3	6		5.77	2.85	6.90	3.27
45 - 78	3	12		11.20	3.92	12.63	4.45
50 - 80	3	11		10.54	4.14	12.27	4.42
55 - 79	3	9		8.58	3.56	10.11	4.03
60 - 78	3	7		6.95	3.06	8.09	3.44

REFERENCES

1. Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.

2. Kuntz KM, Weinstein MC. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2001.

3. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach. Journal of Health Services Research and Policy 2004;9(2):110-8.

4. Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. Journal of the Operational Research Society 2007;58(2):168-76.

5. Davies R, Roderick P, Raftery J. The evaluation of disease prevention and treatment using simulation models. European Journal of Operational Research 2003;150(1):53-66.

6. Kim S-Y, Goldie SJ. Cost-effectiveness analyses of vaccination programmes: A focused review of modelling approaches. PharmacoEconomics 2008;26(3):191-215.

7. Kim JJ, Brisson M, Edmunds WJ, Goldie SJ. Modeling Cervical Cancer Prevention in Developed Countries. Vaccine 2008;26(S10):K76-K86.

8. Habbema JDF, Boer R, Barendregt JJ. Chronic Disease Modeling. In: Heggenhougen K, editor. International Encyclopedia of Public Health. Oxford: Academic Press; 2008. p. 704-9.

9. Torrance GW, Siegel JE, Luce BR. Framing and Designing the Cost-Effectiveness Analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 54-81.

10. Sculpher M, Fenwick E, Claxton K. Assessing Quality in Decision Analytic Cost-Effectiveness Models: A Suggested Framework and Example of Application. PharmacoEconomics 2000;17(5):461-77.

11. Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Medical Decision Making 2007;27(4):491-9.

12. Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? European Journal of Health Economics 2003;4(3):143-150.

13. Mandelblatt J, Fryback D, Weinstein M, Russell L, Gold M. Assessing the effectiveness of health interventions for cost-effectiveness analysis. Journal of General Internal Medicine 1997;12(9):551-8.

14. Ethgen O, Standaert B. Population- versus Cohort-Based Modelling Approaches. PharmacoEconomics 2012;30(3):171-81.

15. Mauskopf J. Prevalence-based economic evaluation. Value in Health 1998;1(4):251-9.

16. Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Medical Decision Making 2010;30(4):426-37.

17. Wilschut JA, Hol L, Dekker E, Jansen JB, van Leerdam ME, Lansdorp-Vogelaar I, et al. Cost-effectiveness Analysis of a Quantitative Immunochemical Test for Colorectal Cancer Screening. Gastroenterology 2011;141(5):1648-55.

18. Loeve F, Brown ML, Boer R, van Ballegooijen M, van Oortmarssen GJ, Habbema JDF. Endoscopic Colorectal Cancer Screening: a Cost-Saving Analysis. Journal of the National Cancer Institute 2000;92(7):557-63.

19. Ramsey SD, Wilschut J, Boer R, van Ballegooijen M. A decision-analytic evaluation of the cost-effectiveness of family history-based colorectal cancer screening programs. Americal Journal of Gastroenterology 2010;105(8):1861-9.

20. de Gelder R, Bulliard JL, de Wolf C, Fracheboud J, Draisma G, Schopper D, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. European Journal of Cancer 2009;45(1):127-38.

21. Boer R, de Koning HJ, Threlfall A, Warmerdam P, Street A, Friedman E, et al. Cost effectiveness of shortening screening interval or extending age range of NHS breast screening programme: computer simulation study. BMJ 1998;317(7155):376-9.

22. de Koning HJ, Martin van Ineveld B, van Oortmarssen GJ, de Haes JCJM, Collette HJA, Hendriks JHCL, et al. Breast cancer screening and cost-effectiveness; Policy alternatives, quality of life considerations and the possible impact of uncertain factors. International Journal of Cancer 1991;49(4):531-7.

23. Groenewoud J, Otten J, Fracheboud J, Draisma G, van Ineveld B, Holland R, et al. Cost-effectiveness of different reading and referral strategies in mammography screening in the Netherlands. Breast Cancer Research and Treatment 2007;102(2):211-8.

24. Van Der Maas PJ, De Koning HJ, Van Ineveld BM, Van Oortmarssen GJ, Habbema JDF, Lubbe KTN, et al. The cost-effectiveness of breast cancer screening. International Journal of Cancer 1989;43(6):1055-60.

25. Koopmanschap MA, Lubbe KTN, van Oortmarssen GJ, van Agt HMA, van Ballegooijen M, Habbema JDF. Economic aspects of cervical cancer screening. Social Science and Medicine 1990;30(10):1081-7.

26. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JDF. Cost-effectiveness of cervical cancer screening: Comparison of screening policies. Journal of the National Cancer Institute 2002;94(3):193-204.

27. Gyrd-Hansen D, Hølund B, Andersen P. A cost-effectiveness analysis of cervical cancer screening: health policy implications. Health Policy 1995;34(1):35-51.

28. Szeto KL, Devlin NJ. The cost-effectiveness of mammography screening: evidence from a microsimulation model for New Zealand. Health Policy 1996;38(2):101-15.

29. Brown ML. Economic Considerations in Breast Cancer Screening of Older Women. The Journals of Gerontology 1992;47(Special Issue):51-8.

30. Lejeune C, Arveux P, Dancourt V, Béjean S, Bonithon-Kopp C, Faivre J. Cost-effectiveness analysis of fecal occult blood screening for colorectal cancer. International Journal of Technology Assessment in Health Care 2004;20(04):434-9.

31. Lejeune C, Dancourt V, Arveux P, Bonithon-Kopp C, Faivre J. Cost-effectiveness of screening for colorectal cancer in France using a guaiac test versus an immunochemical test. International Journal of Technology Assessment in Health Care 2010;26(1):40-7.

32. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective Cost-effectiveness Analysis of Screening Mammography. Journal of the National Cancer Institute 2006;98(11):774-82.

33. Stinnett AA, Mullahy J. Net Health Benefits. Medical Decision Making 1998;18(2):S68-S80.

34. Loeve F, Boer R, van Ballegooijen M, van Oortmarssen G, Habbema J. Final report MISCAN-colon microsimulation model for colorectal cancer: report to the National Cancer Institute Project No. NO1-CN55186. Department of Public Health, Erasmus University; 1998.

35. Loeve F, Boer R, van Oortmarssen G, J,, van Ballegooijen M, Habbema J, D, F,. The MISCAN-colon simulation model for the evaluation of colorectal cancer screening. 1999;32:13–33. Computers and Biomedical Research 1999;32(1):13-33.

36. van Ballegooijen M, Boer R, Habbema JDF, Loeve F, van Oortmarssen GJ, Vogelaar I, et al. Sloan-Kettering Institute for Cancer Research - MISCAN Colon Model Description. 2008. Available from: https://cisnet.flexkb.net/mp/pub/cisnet_colorectal_sloankettering_profile.pdf .

37. O'Mahony JF, van Rosmalen J, Zauber AG, van Ballegooijen M. Multicohort Models in Cost-Effectiveness Analysis: Why Aggregating Estimates over Multiple Cohorts Can Hide Useful Information. Journal of Medical Decision Making 2013;33(3): 407-414.

Paper 4.

Multi-Cohort Models in Cost-Effectiveness Analysis: Why Aggregating Estimates over Multiple Cohorts Can Hide Useful Information

James F. O'Mahony, MA[1,2],
Joost van Rosmalen, PhD[2],
Ann G. Zauber, PhD[3],
Marjolein van Ballegooijen, MD, PhD[2].

1. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland;
2. Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands;
3. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Centre, New York, USA.

Medical Decision Making 33(3).

ABSTRACT

Background: Models used in cost-effectiveness analysis (CEA) of screening programmes may include one or many birth-cohorts of patients. As many screening programmes involve multiple screens over many years for each birth-cohort, the actual implementation of screening often involves multiple concurrent recipient cohorts. Consequently, some advocate modelling all recipient cohorts rather than one birth cohort, arguing it more accurately represents actual implementation. However, reporting the cost-effectiveness estimates for multiple cohorts on aggregate rather than per cohort will fail to account for any heterogeneity in cost-effectiveness between cohorts. Such heterogeneity may be policy relevant where there is considerable variation in cost-effectiveness between cohorts, as in the case of cancer screening programmes with multiple concurrent recipient birth-cohorts, each at different stages of screening at any one point in time.

Objective: The purpose of this study is to illustrate the potential disadvantages of aggregating cost-effectiveness estimates over multiple cohorts, without first considering the disaggregate estimates.

Analysis: We estimate the cost-effectiveness of two alternative cervical screening tests in a multi-cohort model and compare the aggregated and per-cohort estimates. We find instances where the policy choices suggested by aggregate and per-cohort results differ. We use this example to illustrate a series of potential disadvantages of aggregating CEA estimates over cohorts.

Conclusions: Recent recommendations that CEAs should consider the cost-effectiveness of more than just a single cohort appear justified, but the aggregation of estimates across multiple cohorts into a single estimate does not.

INTRODUCTION

The cost-effectiveness of healthcare interventions can differ between patient cohorts. Differences in cost-effectiveness between cohorts mean a common policy for all may not be optimal. Cost-effectiveness analyses (CEAs) should inform policy makers of heterogeneity in cost-effectiveness enabling separate decisions for different cohorts, if necessary [1]. Indeed, the need to reflect heterogeneity between groups is well recognised in CEA as many analyses report separate cost-effectiveness estimates for separate groups.

Despite awareness of the relevance of cohort heterogeneity, the understanding of its implications for CEA model structure are evidently still incomplete in the particular case of CEAs of screening interventions. Screening CEAs conventionally only model a single birth-cohort of screen recipients [2]. However, because screening programmes typically last many years, introducing or changing a screening programme affects multiple birth-cohorts simultaneously and cost-effectiveness may vary between these cohorts. Consequently, some screening CEAs feature all recipient cohorts and report cost-effectiveness on aggregate over all cohorts [3-7]; indeed, this has been explicitly advocated as more representative of actual implementation [2]. However, modelling all screened cohorts together and reporting cost-effectiveness on aggregate will fail to account for any cohort heterogeneity and does not facilitate separate decisions for specific cohorts. The purpose of this study is to illustrate the potential disadvantages of aggregating cost-effectiveness estimates over multiple cohorts, without first considering the disaggregate estimates. This study is intended to inform both guidance on appropriate model structures and the correct interpretation of results from multiple cohort models.

This study uses the specific context of screening to address the question of aggregation across cohorts for two reasons. Firstly, the existing screening CEAs that model multiple cohorts and report cost-effectiveness on aggregate show that this approach is used in current practice. Secondly, the methodological research advocating this approach was motivated by the fact that screening typically involves multiple concurrent recipient birth-cohorts [2]. Although our study most

immediately applies to screening, the questions around aggregation apply more broadly to any example where cost-effectiveness varies between patient groups.

This paper is structured as follows. The following section provides the background to this study, giving an overview of different model structures and terminology, and briefly reviewing how cohort heterogeneity has been addressed in the literature. The analysis section describes the methods and results of a CEA of cervical cancer screening and explains four potential problems of aggregating cost-effectiveness estimates, using the cervical screening CEA as an example. The discussion considers the implications of these issues for decision making and the potential trade-offs between single and multiple cohort modelling.

Background

Model Types and Terminology

Cohorts can be defined either as groups starting an intervention at a common point in time or of a common birth-year. The birth-year definition is appropriate in the case of screening, as screening eligibility is typically determined by age and because the cost-effectiveness of screening can vary with age. We categorise cohorts by adapting the terminology used by Hoyle and Anderson of prevalent and incident cohorts [8]. The terms do not correspond to any disease state in this context, but to screening eligibility. The single birth-cohort starting screening at the programme start age in the current period is the current incident cohort; cohorts already within the screening age range when a programme is introduced or changed are the prevalent cohorts; and, cohorts currently younger than the programme start age are the future incident cohorts. Note that there can be many prevalent and future incident cohorts, each of different birth years. We collectively describe the prevalent cohorts and the current incident cohort as the current cohorts, while the future incident cohorts are simply described as the future cohorts.

Cost-effectiveness may vary between prevalent cohorts at any point in time due to differences in age and screening histories. Consequently, even if unit costs and treatment effectiveness are constant over time, screening interventions can lead to

cohort heterogeneity in cost-effectiveness. The birth-year cohorts in this analysis are mutually exclusive, however the cohort groups used in published CEAs are not always mutually exclusive in that one cohort group can be a subset of another broader cohort composed of more than one birth-year cohort.

Cohort Heterogeneity in the Literature

Single cohort models are widely used in CEA [8-10]. They model the costs and effects of an intervention for one cohort of patients. In the case of CEAs of screening, this cohort is typically a single birth-year cohort [11-14]. The most widely recognised rationale for multi-cohort modelling is in the context of infectious diseases, where multiple cohorts are necessary to simulate herd immunity [15, 16]. However, the rationale considered in this study is the need to capture cohort heterogeneity. A single cohort model is sufficient if that cohort's cost-effectiveness is representative of all recipient cohorts. Conversely, if cost-effectiveness is anticipated to vary significantly between cohorts, then a multi-cohort model will be more appropriate [8, 17, 18].

The literature addressing the cohort heterogeneity rationale for multi-cohort modelling is sparse. However, two previously published papers in this journal have directly addressed the topic and advocate multi-cohort models as more representative of the actual healthcare implementation [2, 8]. Dewilde and Anderson present a CEA of cervical screening, showing its average cost-effectiveness ratio (ACER) to be considerably higher when prevalent cohorts are considered in addition to the incident cohort alone [2]. More recently, Hoyle and Anderson recognised that if cost-effectiveness varies between prevalent and incident cohorts, then the number of future incident cohorts modelled will influence the aggregate cost-effectiveness estimate [8]. They also considered the possibility that cost-effectiveness may vary between incident cohorts. Consequently, they recommend CEAs include all current and all future cohorts and be reported as a "combined cohorts ICER". However, an accompanying editorial by Kuntz et al. questioned whether it is appropriate to aggregate results over multiple cohorts [19]. Similarly, Karnon et al. noted that Dewilde and Anderson's multi-cohort approach can be used to account for the effect of cohort heterogeneity on aggregate cost-effectiveness, but

suggested separate per-cohort analyses be used where interventions can be applied differently to separate cohorts [18].

Examples of multiple cohort models from the screening literature include models featuring the current cohorts only [3, 4, 7], while others also model future incident cohorts [5, 6]. Notably, cost-effectiveness is reported on aggregate over all cohorts in all these examples.

ANALYSIS

Methods

We present a CEA of two alternative cervical cancer screening tests as an illustrative example of the consequences of aggregating cost-effectiveness estimates over cohorts. We use the MISCAN-cervix microsimulation model developed by the Department of Public Health at Erasmus MC, the Netherlands. The model simulates the individual life histories of women from birth until death. The model generates age and time specific cancer incidence and mortality estimates. Alternative screening scenarios can be simulated in the model, whereby screening may detect disease in the preclinical phase, permitting early treatment to prevent further disease development and subsequent death from the cancer. Quality of life weights are applied to the different health states to estimate health effects of screening in terms of quality adjusted life years (QALYs).

The MISCAN-cervix model including its parameter values and data sources can be found in previous publications [20-23]. The rates of disease incidence and the sojourn times used in the model are derived from an age-period-cohort model of cervical cancer in the years prior to the introduction of population screening in the Netherlands [24]. The data from the age-period-cohort study are supplemented by calibrating MISCAN-cervix to the observed rates of cervical intraepithelial neoplasia within the screened population between 1997 and 2001 since the introduction of population screening [25] and the observed age-specific rates of infections with high risk HPV types [26].

The model also includes vaccination against HPV types 16 and 18; precursors to cervical cancer. The characteristics of the vaccine are as modelled in a recent CEA of the bivalent vaccine in the Netherlands [21]. The vaccine is administered in three doses and is assumed equally effective in all vaccinated cohorts. The vaccine coverage rate simulated is 85% and there is no selection effect assumed regarding vaccine coverage and the risk of infection. Vaccine efficacy is assumed to be 70% against cancer, 35 % against pre-invasive lesions and 1.5% against HPV infections. The protection of the vaccine is assumed to be lifelong in this analysis.

We compare two screening scenarios. The baseline scenario is the current Dutch cervical screening programme of 7 lifetime screens between the ages of 30 and 60 at 5 year intervals, using a primary cytology screen, with cytology triage for abnormal primary smears. The alternative scenario is a change in 2012 to screening with a primary human papillomavirus (HPV) DNA test, followed by three cytology triage tests for abnormal primary screen results. The screening age range and frequency remain unchanged.

The model simulates 46 separate birth-year cohorts. The size of each cohort is matched to the number of female live births in each respective birth year in the Netherlands. The current incident cohort starting screening in 2012 is aged 30. There are 30 prevalent birth-cohorts aged 31 to 60 in 2012. Those in the prevalent cohorts will have already experienced at least one screening round prior to 2012 using the current test. The 15 future incident cohorts that have not yet started screening are aged between 29 and 15 in 2012. Vaccination against HPV types 16 and 18 was introduced in the Netherlands in 2009 for girls aged 12 with catch-up for those aged up to and including 16. Consequently, the youngest four future incident cohorts in the model have been offered vaccination.

All costs and health effects due to the screening programmes are included in the model from 2012 until death for each cohort. Costs and effects are discounted at 4% to the discount year of 2012. The costs used are from 2008 and are reported in euro (1 euro = 1.45 U.S. dollars; 6th July 2011). We assume the unit costs of screening, follow up and avoided treatment remain constant over time. The parameters determining health gain from the early detection of cancer and precancerous lesions

are also assumed constant over time. Age specific disease incidence is also assumed to remain the same, except for those vaccinated cohorts, where disease incidence falls in accordance with the vaccine's expected effectiveness.

The costs and effects are estimated both on a disaggregated per-cohort basis and an aggregated basis over multiple cohorts. These results are presented as the average cost-effectiveness ratio (ACER) of each intervention compared to no screening and the incremental cost-effectiveness ratio (ICER) of switching from the current primary cytology test to the HPV test. Both ACERs and ICERs are reported in terms of euro per QALY.

No direct funding was provided for this project. The first author is supported by the Health Research Board of Ireland and by the US National Cancer Institute under grant U01 CA115953. Neither funding body had a role in the study.

Results

Figure 1 shows the ACERs of both tests for each cohort and on aggregate. Figure 2 shows the ICERs for HPV screening relative to cytology screening. Table 1 reports the effects, cost, ACERs and ICERs for selected cohorts and the multi-cohort aggregates. The table also reports the proportion of total net costs of screening for the selected cohorts, using the proportion of total costs under primary cytology as an indication of the relative weight of each cohort in the aggregate.

The per-cohort ACERs and ICERs fall into groups of cohorts with common screening histories and vaccination status. For example, in 2012 the five oldest prevalent cohorts, -30 to -26, all have only one lifetime screening left at age 60, while cohorts -25 to -21 have two left, one at 55 and one at 60. The common cost-effectiveness within each group reflects the assumption of constant costs and effects: the small remaining variation between cohorts is a consequence of the stochastic nature of the simulation model.

The per-cohort ACERs show both screening tests to be less cost-effective in vaccinated cohorts relative to unvaccinated incident cohorts and increasingly less cost-effective with age in the prevalent cohorts. HPV testing is more costly and

more effective than cytology in every cohort. Cytology is weakly dominated by a combination of HPV testing and no intervention in cohorts -30 to -16, the oldest 15 prevalent cohorts, so no ICERs of HPV testing are reported for them. The per-cohort ICER of HPV relative to primary cytology screening is essentially uniform among the unvaccinated incident cohorts (0 to 11), at approximately €47,700 per QALY. The HPV test is more cost-effective in prevalent cohorts -10 to -1 than in the incident cohorts, being the most cost-effective in the youngest 5 prevalent cohorts with and ICER of €31,800 per QALY. HPV testing is markedly less cost-effective in the vaccinated future cohorts 12 to 15, with ICERs in excess of €500,000 per QALY. The aggregate ICER of HPV testing over the current cohorts is approximately €38,300 per QALY, while the aggregate ICER all cohorts is somewhat higher at €45,400.

The results show how the cost-effectiveness of interventions can vary significantly between cohorts. The markedly high ACERs for both screening methods in the vaccinated cohorts are a consequence of the simulated reduction in disease incidence following vaccination. Similarly, the rising ACERs with age within the prevalent cohorts are attributed to the fact that both disease incidence and the potential life years that can be gained fall with age. The changing relative cost-effectiveness of the two tests is attributed to the fact that while HPV testing is more effective due to higher sensitivity, it can lead to over-detection of transient HPV infections and non-cancerous lesions, which are more common in younger women.

Figure 1. Per-cohort and aggregate average cost-effectiveness ratios of cytology and HPV testing compared to no screening. The prevalent cohorts are numbered -30 to -1 starting with the oldest cohort, 0 for the current incident cohort and from 1 to 15 for the future incident cohorts. The ACERs of cytology and the HPV test are in grey and black respectively, with dashes for each individual cohort, solid lines for all cohorts on aggregate and dotted lines for current cohorts on aggregate.

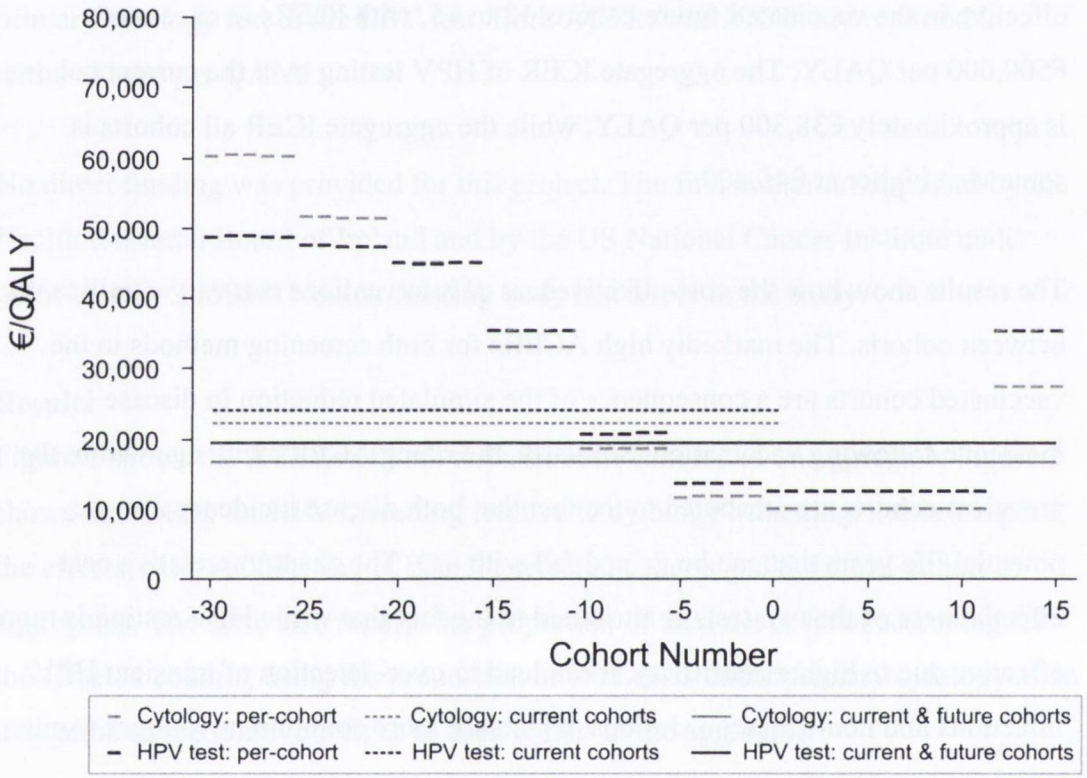


Figure 2. Per-cohort and aggregate incremental cost-effectiveness ratios of HPV testing compared to cytology. The ICER of HPV testing relative to cytology for those cohorts in which cytology is not weakly dominated by a combination of HPV testing and no screening is shown with dashes. The aggregate ICERs over all cohorts and the current cohorts are shown with the solid and dotted lines respectively.

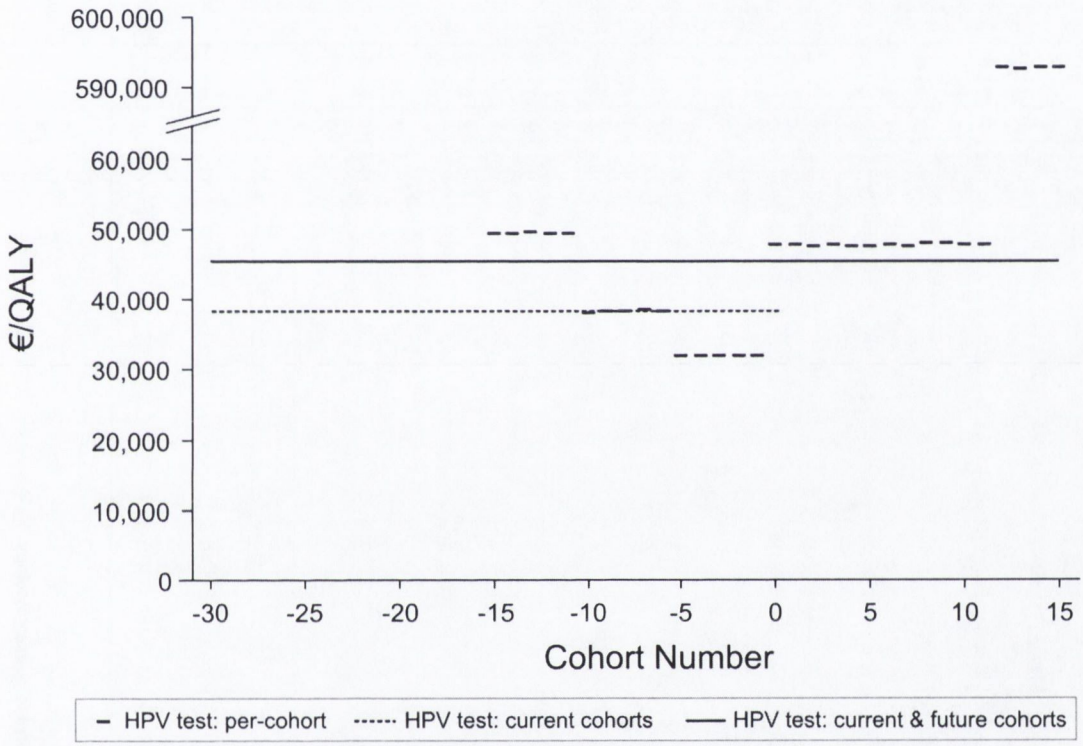


Table 1. Discounted costs, effects, average cost-effectiveness ratios of primary cytology and primary HPV based screening compared to no screening and incremental cost-effectiveness ratios of primary HPV testing compared to primary cytology for selected cohorts and multi cohort aggregates

	Cytology primary test screening				HPV primary test screening			
Cohort	Effects, QALYs 000s*	Costs, €Millions†	ACER,‡ €/QALY	Proportion of total costs, %	Effects, QALYs 000s	Costs, €Millions†	ACER,‡ €/QALY	ICER,§ €/QALY
-30	38	2,296	60,200	0.8	58	2,800	48,700	N/A
-25	84	4,347	51,600	1.6	112	5,325	47,400	N/A
-20	144	6,469	45,000	2.4	177	7,954	44,900	N/A
-15	237	7,806	33,000	2.9	274	9,645	35,200	49,400
-10	433	8,051	18,600	3.0	486	10,039	20,700	37,900
-5	611	7,110	11,600	2.6	672	9,036	13,500	31,800
0	795	7,917	10,000	2.9	851	10,602	12,500	47,700
5	709	7,049	9,900	2.6	759	9,439	12,400	47,500
10	614	6,115	10,000	2.3	658	8,188	12,500	47,700
15	205	5,623	27,400	2.1	208	7,326	35,200	592,800
-30 to 0	7,797	171,858	22,000	63.4	8,903	214,158	24,100	38,300
1 to 15	8,421	99,085	11,800	36.6	8,968	131,812	14,700	59,800
-30 to 15	16,218	270,943	16,700	100.0	17,871	345,970	19,400	45,400

* Quality adjusted life years
† Costs in 2008 prices.
‡ Average cost-effectiveness ratio compared to no screening
§ Incremental cost-effectiveness ratio of primary HPV testing compared to primary cytology

Four Problems of Aggregation

Both Dewilde and Anderson and Hoyle and Anderson call attention to the fact that many CEA models do not accurately represent the policy choices they are to inform: while CEAs typically only assess cost-effectiveness for a single incident cohort, the policy decisions they inform are almost never for that cohort alone, but also for other cohorts, both in the present and the future [2, 8]. Consequently, Hoyle and Anderson suggest that CEAs include all present and future recipient cohorts. Given the need to capture cohort heterogeneity, this suggestion appears sensible. However, reporting cost-effectiveness as a single aggregate estimate for all cohorts seems inappropriate for the following reasons: (i) aggregate estimates over many cohorts may hide useful information from decision makers; (ii) the choice of which cohorts to include appears logically problematic; (iii) aggregate estimates demand significant assumptions about the future, which results in large uncertainty in aggregate estimates; (iv) aggregate modelling prompts broader questions about decision making over multiple periods for healthcare priority setting. This section explains each of these problems in detail.

Aggregate Results Hide Differences

Reporting an aggregate cost-effectiveness estimate across all recipient cohorts carries an implicit assumption that a common policy decision will be taken for all cohorts. However, different reimbursement decisions can clearly be made for different cohorts in many cases. Where selective reimbursement is possible, decision makers would be better served by cost-effectiveness estimates disaggregated per-cohort, allowing them to approve the intervention for those cohorts identified as cost-effective to treat and withhold it from those which are not.

Our example shows how aggregate results can lead to inappropriate policy choices in terms of cost-effectiveness. If the cost-effectiveness threshold was €40,000 per QALY, the aggregate ICER over the current cohorts of €38,300 per QALY would indicate the HPV test is cost-effective. However, according to the per-cohort estimates, HPV screening would not be cost-effective in any incident cohort or the oldest 20 of the 30 prevalent cohorts. Conversely, if the aggregate ICER over all cohorts of €45,400 per QALY was used, then the HPV test would be deemed not

cost-effective, but the per-cohort estimates indicate that it would be cost-effective for prevalent cohorts -10 to -1. Similarly, relying on the single incident cohort ICER of €47,700 per QALY would also lead to a rejection of the test for all cohorts.

There are certain cases where disaggregated per-cohort estimates are not appropriate. These include interventions where the selective allocation of the intervention is not possible, such as water fluoridation. Aggregate estimates are also appropriate in cases where the effects of an intervention are shared across cohorts, such as herd immunity from vaccination. Another consideration is that it may be impractical to offer different interventions to different cohorts, as the costs of tailoring the intervention to each cohorts may outweigh the benefits. For example, in the case of screening, offering different screening strategies to different birth-cohorts may reduce adherence, possibly leading to a greater health loss than would be gained by optimising the programme for each cohort.

Logical Problems of Including and Excluding Cohorts

A related problem of aggregation is that results depend on which cohorts are modelled. There may be many potential recipient cohorts, but not all may be cost-effective to treat. We included vaccinated women as an example of cohorts in which the HPV test has poor incremental cost-effectiveness. Presumably these cohorts should be excluded when estimating aggregate cost-effectiveness of HPV primary screening, as they are unlikely to receive that strategy due to its poor cost-effectiveness. However, using cost-effectiveness as an inclusion criterion for cohorts in the model leads to a certain circularity: if only cost-effective cohorts are included in the analysis (assuming at least one cost-effective cohort exists), then the intervention will necessarily be cost-effective on aggregate for the selected cohorts.

An analogous problem arises if cost-effectiveness is not an inclusion criterion: if all cohorts have ICERs that are either all above or all below the threshold, then including some or all cohorts will make no difference to the reimbursement decision. However, if some cohorts' ICERs are above the threshold and others' are below it, then an aggregate ICER, which must be either above or below the threshold, will necessarily imply an incorrect decision for some cohorts.

Consequently, the aggregate multi-cohort approach could be interpreted as either not enhancing decision making or necessarily leading to errors.

Requirement of Additional Assumptions and the Impact on Uncertainty

Hoyle and Anderson's recommendation that multi-cohort models include all present and future recipient cohorts means CEA would then encompass the entire expected implementation lifetime of an intervention. Modelling all recipient cohorts requires considerable assumptions, especially regarding future cohorts. A common assumption within CEA is that unit costs and health effects remain constant over time [27]. However, as noted above, even with the assumption of constant costs and effects, cost-effectiveness can vary between prevalent and incident cohorts; in which case, the number and size of the future incident cohorts modelled will influence aggregate cost-effectiveness estimates.

While in our example the advent of vaccination means there will be a predictable date by which screening is likely to change, most interventions do not have accurately predictable lifetimes. In a recent study, Hoyle uses evidence of the lifetime and volume of pharmaceutical interventions in the past to inform assumptions regarding the implementation of new interventions and adjust the cost-effectiveness estimates accordingly [27]. However, the validity and reliability of past evidence of previous interventions as a predictor for new interventions in the future are certainly dubious. Consequently, the cost-effectiveness estimates of aggregate models may seem arbitrary in part, given their reliance on assumptions regarding highly uncertain future use.

Modelling cost-effectiveness over an intervention's entire implementation lifetime significantly increases the scope for uncertainty of estimates. Such additional uncertainty is unwelcome. Uncertainty in input parameters will grow the further into the future they are projected. While this uncertainty also applies to disaggregate multi-cohort models that report cost-effectiveness separately for each cohort, it is limited, to a degree, to the cost-effectiveness estimates of future incident cohorts. Given that it will be possible to review the cost-effectiveness evidence for most interventions at a later date and revise the reimbursement decision if necessary, the cost-effectiveness of the current and near future cohorts is more relevant to the

policy choices currently facing decision makers than that of cohorts far in the future. Consequently, modelling an intervention over its entire expected implementation lifetime unnecessarily adds uncertainty to current cost-effectiveness estimates.

Decision Making over Multiple Periods

Cohort heterogeneity arises in our example due to differences in age and screening history at the time of the policy change and vaccination status. Cohort heterogeneity could also occur if costs or effects change over time or if differential discounting of costs and health effects is applied, as is required in Belgium, the Netherlands and Poland [28-30] . Hoyle and Hoyle and Anderson use the examples of falling drug prices and differential discounting respectively as rationales for modelling intervention' cost-effectiveness over their entire expected lifetimes [8, 27].

Falling drug prices or the application of differential discounting can lead to inconsistencies between the per-cohort and lifetime analyses of an intervention's cost-effectiveness. For example, a drug may not be cost-effective at its current patent protected price, but may be cost-effective if sufficient future cohorts enjoying a lower post-patent price are added to a CEA [27]. Conversely, post-patent price reductions of comparator drugs can lead to interventions becoming not cost-effective when analyses are extended to include future periods [31]. Similarly, as cost-effectiveness improves with the inclusion of more future cohorts under differential discounting [32], the cost-effectiveness rank order of two interventions may switch when compared first on a per-cohort basis and then over their implementation lifetimes, where those lifetimes are unequal. These examples of inconsistencies between the per-cohort and the lifetime perspectives prompt the question which is more appropriate; especially as health priority setting is a repeated resource allocation problem reoccurring each year, not a once-off decision over a finite horizon.

DISCUSSION

Determining the most appropriate model structure requires an awareness of both the actual policy question faced by decision makers and any assumptions implicit in the model structure and the presentation of results. Relying on single cohort models can be interpreted as embodying an assumption that all cohorts will exhibit the same cost-effectiveness as the current incident cohort. However, recommending aggregate reporting of cost-effectiveness estimates over all current and future cohorts could equally be interpreted as implicitly assuming that interventions cannot be selectively reimbursed and that the lifetime of their use is the appropriate basis for comparisons to other interventions. Consequently, rather than estimating cost-effectiveness over all recipients as suggested by Dewilde and Anderson and Hoyle and Anderson, we support Karnon et al.'s suggestion that results be reported on a disaggregate basis [18], except where there is a reasonable rationale to report aggregate estimates, as described above. Reporting the results of multi-cohort models on a per-cohort basis is equivalent to using multiple single-cohort models. However, this is simply a semantic distinction; what matters is the useful reporting of results to decision makers.

Regarding the useful communication of cost-effectiveness estimates to decision makers, this study reported both ACERs and ICERs of the various strategies. The use of ACERs was to highlight the heterogeneity of cost-effectiveness between cohorts. However, in practice, reimbursement decisions should be made using ICERs not ACERs.

If per-cohort and aggregate optimised strategies differ, then the aggregate strategy will be suboptimal, given the assumption separately specified interventions can be implemented without any additional costs. However, where it is costly to provide per-cohort optimised strategies, the trade-off between costs and benefits of per-cohort optimisation should be considered.

Providing disaggregated per-cohort estimates may demand more work, as estimates have to be made for each cohort separately. Given that the principal modelling effort is the gathering of parameter estimates and model specification, the additional

effort of generating disaggregated estimates should be relatively small. However, a remaining drawback for decision makers is the potential difficulty of interpreting multiple estimates. Given the difficulty decision makers may have interpreting multiple cost-effectiveness estimates, cost-effectiveness analysts may to consider carefully how best to communicate their results. A pragmatic approach might be to present the efficient strategies for the incident cohort as the base case and only present cost-effectiveness estimates for other cohorts in cases in which differences in cost-effectiveness would imply a meaningful difference in the optimal intervention strategy.

CONCLUSION

The recommendations of the current literature on multi-cohort modelling that CEAs should include not only the current incident cohort, but also the current prevalent and future incident cohorts seem sound. However, the aggregation of cost-effectiveness estimates for all cohorts into a single estimate does not appear useful to the actual choices faced by decision makers in most cases. Therefore, we suggest consideration of the cost-effectiveness of more than just a single incident cohort in cases in which cost-effectiveness is likely to vary between cohorts, with estimates reported on a disaggregate per-cohort basis in addition to the overall estimate for all cohorts. This applies to CEAs in general, but is particularly relevant for analyses of screening.

REFERENCES

1. Torrance G, Siegel J, Luce B. Framing and Desiging the Cost-Effectiveness Analysis. In: Gold M, Siegel J, Russell L, Weinstein M, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 54-81.

2. Dewilde S, Anderson R. The cost-effectiveness of screening programs using single and multiple birth cohort simulations: A comparison using a model of cervical cancer. Med Decis Making. 2004;24(5):486-92.

3. Berchi C, Bouvier V, Réaud J-M, Launoy G. Cost-effectiveness analysis of two strategies for mass screening for colorectal cancer in France. Health Econ. 2004;13(3):227-38.

4. Maxwell GL, Carlson JW, Ochoa M, Krivak T, Rose GS, Myers ER. Costs and effectiveness of alternative strategies for cervical cancer screening in military beneficiaries. Obstet Gynecol. 2002;100(4):740-8.

5. Ramsey SD, Wilschut J, Boer R, van Ballegooijen M. A decision-analytic evaluation of the cost-effectiveness of family history-based colorectal cancer screening programs. Am J Gastroenterol. 2010;105(8):1861-9.

6. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective Cost-effectiveness Analysis of Screening Mammography. J Natl Cancer Inst. 2006;98(11):774-82.

7. Tosteson ANA, Stout NK, Fryback DG, Acharyya S, Herman BA, Hannah LG, et al. Cost-effectiveness of digital mammography breast cancer screening. Ann Intern Med. 2008;148(1):1-10.

8. Hoyle M, Anderson R. Whose costs and benefits? Why economic evaluations should simulate both prevalent and all future incident patient cohorts. Med Decis Making. 2010;30(4).

9. Davies R, Roderick P, Raftery J. The evaluation of disease prevention and treatment using simulation models. Eur J Oper ResS. 2003;150(1):53-66.

10. Mauskopf J. Prevalence-based economic evaluation. Value Health. 1998;1(4):251-9.

11. Ahern CH, Shen Y. Cost-effectiveness analysis of mammography and clinical breast examination strategies: A comparison with current guidelines. Cancer Epidemiol Biomarkers Prev. 2009;18(3):718-25.

12. Anderson R, Haas M, Shanahan M. The cost-effectiveness of cervical screening in Australia: what is the impact of screening at different intervals or over a different age range? Aust N Z J Public Health. 2008;32(1):43-52.

13. Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ. Cost-effectiveness of cervical cancer screening with Human Papillomavirus DNA testing and HPV-16,18 Vaccination. J Natl Cancer Inst. 2008;100(5):308-20.

14. Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, Wilschut JA, Zauber AG, van Ballegooijen M. Stool DNA testing to screen for colorectal cancer in the Medicare population. Ann Intern Med. 2010;153(6):368-77.

15. Brisson M, Edmunds WJ. Economic evaluation of vaccination programs: The impact of herd-immunity. Health Econ. 2003;23(1):76-82.

16. Edmunds WJ, Medley GF, Nokes DJ. Evaluating the cost-effectiveness of vaccination programmes: a dynamic perspective. Stat Med. 1999;18(23):3263-82.

17. Feuer EJ. Chapter 1: Modeling the impact of adjuvant therapy and screening mammography on U.S. breast cancer mortality between 1975 and 2000: Introduction to the problem. J Natl Cancer Inst Monogr. 2006;2006(36):2-6.

18. Karnon J, Brennan A, Akehurst R. A critique and impact analysis of decision modeling assumptions. Med Decis Making. 2007;27(4):491-9.

19. Kuntz KM, Fenwick E, Briggs A. Appropriate cohorts for cost-effectiveness analysis: To mix or not to mix? Med Decis Making. 2010;30(4):424-5.

20. Habbema JDF, van Oortmarssen GJ, Lubbe JTN, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.

21. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of Human Papillomavirus vaccination in the Netherlands. J Natl Cancer Inst. 2009;101(15):1083-92.

22. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JDF. Cost-effectiveness of cervical cancer screening: Comparison of screening policies. J Natl Cancer Inst. 2002;94(3):193-204.

23. de Kok IMCM, van Rosmalen J, Sasieni P, Dillner J, Iftner T, van Ballegooijen M. Cost-effectiveness of primary hpv screening compared to primary cytology screening for cervical cancer in Europe. In submission.

24. van Ballegooijen M. Effects and costs of cervical cancer screening. Rotterdam: Erasmus University; 1998.

25. Nederlandse Vereniging voor Pathologie (Dutch Society for Pathology). PALGA: The Nationwide Network and Registry of Histo- and Cytopathology in the Netherlands.

26. Jacobs MV, Walboomers JMM, Snijders PJF, Voorhorst FJ, Verheijen RHM, Fransen-Daalmeijer N, et al. Distribution of 37 mucosotropic HPV types in women with cytologically normal cervical smears: The age-related patterns for high-risk and low-risk types. Int J Cancer. 2000;87(2):221-7.

27. Hoyle M. Accounting for the drug life cycle and future drug prices in cost-effectiveness analysis. Pharmacoeconomics. 2011;29(1):1-15.

28. Cleemput I, van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for Pharmacoeconomic Evaluations in Belgium. Brussels: Health Care Knowledge Centre (KCE), 2008 Contract No.: 78C.

29. Oostenbrink JB, Bouwmans CAM, Koopmanschap MA, Rutten FFH. Guidelines for Pharmacoeconomic Research, Updated Version. Diemen: College voor zorgverzekeringen, 2004.

30. Król Z. Guidelines for conducting Health Technology Assessment (HTA). Warsaw: Agency for Health Technology Assessment, 2009.

31. Shih Y-CT, Han S, Cantor SB. Impact of Generic Drug Entry on Cost-Effectiveness Analysis. Med Decis Making. 2005;25(1):71-80.

32. O'Mahony JF, de Kok IMCM, van Rosmalen J, Habbema JDF, Brouwer W, van Ballegooijen M. Practical implications of differential discounting in cost-effectiveness analyses with varying numbers of cohorts. Value Health. 2011;14(4):438-42.

Paper 5.

The influence of disease risk on the optimal time interval between screens for the early detection of cancer: A mathematical approach

James F. O'Mahony, MA[1,2],
Joost van Rosmalen, PhD[1],
Nino.A. Mushkudiani, PhD[3],
Frans-Willem. Goudsmit, MSc[1],
Rene M.J.C. Eijkemans, PhD[4],
Ewout.W. Steyerberg, PhD[1],
J. Dik F. Habbema, PhD[1].

1. Dept of Public Health, Erasmus MC – University Medical Centre Rotterdam, Rotterdam, the Netherlands;
2. Dept of Health Policy and Management, Trinity College Dublin, Dublin, Ireland;
3. Statistics Netherlands, Den Haag, the Netherlands;
4. Dept of Biostatistics, UMC – University Medical Centre Utrecht, Utrecht, the Netherlands;

Under review at Medical Decision Making.

ABSTRACT

The intervals between screens for the early detection of diseases such as breast and colon cancer suggested by screening guidelines are typically based on the average population risk of disease. With the emergence of ever more biomarkers for cancer risk prediction and the development of personalized medicine, there is a need for risk-specific screening intervals. The interval between successive screens should be shorter with increasing cancer risk. A risk-dependent optimal interval that yields the greatest health gain for a given willingness to pay threshold is ideally derived from a cost-effectiveness analysis using a validated simulation model. However, this is time-consuming and costly. We propose a simplified mathematical approach for the exploratory analysis of the implications of risk level on optimal screening interval. We develop a mathematical model of the optimal screening interval for breast cancer screening. Optimality in this case is maximizing the expected health gain for a given willingness to pay for a unit of health. We verified the results by programming the simplified model in the MISCAN-Breast microsimulation model and comparing the results. We validated the results by comparing them with the results of a full, published MISCAN-Breast cost-effectiveness model for a number of different risk levels. The fit of the mathematical model's results to those of MISCAN were satisfactory in both the verification and validation. We conclude that the mathematical approach can be used for giving an indication of the impact of disease risk on the optimal screening interval.

INTRODUCTION

Screening for preclinical disease can improve health outcomes by enabling early treatment of cancer and its precursor lesions. Screening has been proven to improve health outcomes in a number of cancers, including breast cancer [1-4]. Increasing the screening frequency can yield greater health gains, but these are traded-off against the greater costs and risks of more intensive screening, such as false positives, overdiagnosis and repeated x-ray exposure.

Randomized controlled trials have been crucial in establishing the effectiveness of breast, lung and colon cancer screening. These trials assessed various screening age ranges and screening frequencies. However, the trialed screening strategies are not necessarily optimal in terms of cost-effectiveness, in that they many not yield the maximum health gain for a given cost-effectiveness or willingness to pay threshold. It would be impossible to conduct controlled trials to identify optimal strategies given the number of possible alternatives, due to time constraints and ethical considerations [5]. Therefore, simulation models of screening based on the results of published trials have been developed to estimate the effectiveness and cost-effectiveness of alternative screening strategies. These cost-effectiveness analyses (CEA) typically only estimate the optimal age range and interval for screening average-risk populations. Thus, strictly speaking, the optimal policies suggested by CEAs only apply to average-risk individuals.

Those at higher risk of disease have greater potential to gain from more intensive screening than those at average-risk. This is reflected in screening guidelines for those with a family history of breast or colorectal cancer, which recommend screening with a higher frequency and earlier starting age than for the general population [6,7]. However, since CEA estimates are usually unavailable for high-risk groups, screening guidelines for such groups are typically based on expert opinion.

The question of optimal screening intensity is of increased significance with improved knowledge of biomarkers of individual disease risk. These risk markers include genetic and biochemical markers, but also family history and behavioral risk

factors such as smoking, diet and physical activity. Growing knowledge of prognostic markers will lead to risk groups becoming increasingly differentiated, thus expanding the scope for risk-specific screening strategies.

A full cost-effectiveness analysis for each level of disease risk is time-consuming, costly and requires the input of experienced modelers. There is a need for a quick first approximation of what an elevated risk implies for the screening interval. The purpose of this paper is therefore to present an easy-to-apply tool for a first estimation of the optimal screening interval for high-risk groups. This will provide an informed basis for adapting screening guidelines for high-risk groups where specific CEA estimates are not yet available.

Previous Literature

Most current screening models are simulation models, employing cohort or microsimulation approaches. Simulation is useful as it can easily handle complex screening models with many parameters and health states. Ramsey et al. and Knudsen et al. provide useful overviews of simulation models and their application to cancer screening [8,9]. The MISCAN microsimulation model used in this study is an example of such simulation models [10-12].

The cost-effectiveness of screening can also be estimated using mathematical models. These models offer the advantages that they can be analytically solved for optimal solutions and that these solutions can then be rapidly applied in other applications. However, requiring that the solution to a model is mathematically tractable may limit the model's level of realism. Below we give an overview of relevant mathematical models regarding screening programs.

An early example of a mathematical model of optimal screening intervals is Kirch and Klein [13]. They present a relatively simple model of screening, showing that the optimal screening interval varies proportionally to the square root of the incidence rate. This relationship is derived under the simplifying assumptions of perfect screening sensitivity and constant duration of the preclinical period during which a disease is screen-detectable.

Zelen [14] presents a more detailed model of screening that allows for both imperfect sensitivity and a random duration of the preclinical period. The model uses a utility function for the detection of disease weighted by the probability of cure for screen and clinically detected cancer. The optimal screening intervals are found as those maximizing the utility function subject to a constraint on the number of screens, which serves as proxy for a budget constraint. They illustrate their model using the example of breast cancer screening under the assumption of constant disease incidence.

Lee and Zelen [15] extend the model presented by Zelen to consider the possibility that disease incidence varies with age. The measure of screening efficacy used in the study is schedule sensitivity, which is the proportion of total disease cases detected by the screening program over the period of its implementation. The optimal strategy is found using what is described as a threshold method, whereby each screen is scheduled following the initial screen so that the probability of an individual having preclinical disease at each screen is equal to that at the first screen. Lee and Zelen also consider how screening intervals and age ranges might vary in high-incidence groups. They find that the optimal screening schedule for a high-risk group starts at a younger age and has shorter intervals relative to that of the average-risk population.

Baker [5] presents a mathematical model of breast cancer screening based on tumor growth rates from which the probabilities of detection or death are derived. The model finds optimal policies that minimize an objective function of the costs of disease, which includes the costs of screening and life lost. It does not consider how the optimal screening interval changes with disease incidence.

Parmigiani (1993) presents a more general model in which screening sensitivity can vary with age and duration of the preclinical state [16]. The model also describes an objective function containing the costs of screening and the value of health gained. Illustrative applications of this model for breast cancer screening are given in Parmigiani and Kamlet, and Parmigiani (1997) [17,18]. Neither study considered how screening intensities should vary between risk subgroups.

This overview shows that there are already a number of detailed mathematical models of screening effectiveness. However, almost none of these studies have been conducted in the context of conventional CEAs, whereby the ratio of incremental costs to incremental health effects is compared to a cost-effectiveness threshold to identify the optimal screening interval given the threshold willingness to pay. In addition, in most studies, the motivation for investigating optimal screening interval is to understand how the interval might vary with age, rather than between risk subgroups.

Contribution

The purpose of this paper is to present a simple mathematical model to estimate the relationship between disease incidence and the optimal screening interval, accounting for imperfect screening sensitivity and specificity and a random duration of the preclinical disease state.

Our model of the optimal screening interval is more sophisticated than that of Kirch and Klein. While our model is less detailed than that of Parmigiani (1993), it is easier to apply and more accessible in its exposition. It differs from Lee and Zelen, Baker and others that do not consider interval optimization in terms of conventional CEA.

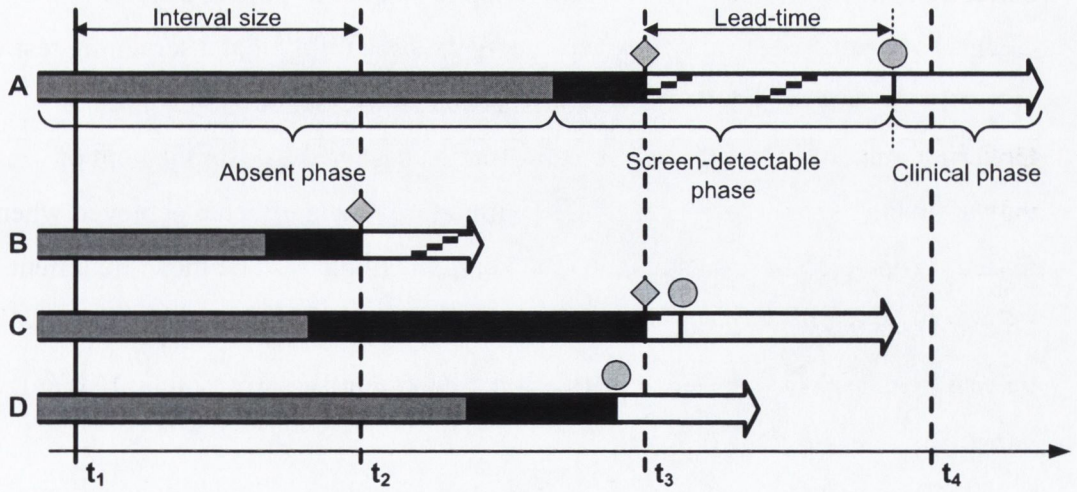
Unlike other studies in the literature, our model has the benefit of being verified and validated against a published microsimulation CEA model, namely the MISCAN-Breast model. The verification is to check that the mathematical model and MISCAN yield the same results when the same simplified assumptions used in the mathematical model are programmed in MISCAN. The validation is to check whether the mathematical model approximates the change in screening intervals with disease incidence as calculated by the full MISCAN microsimulation model: i.e., whether our relatively simple mathematical model adequately captures the underlying relationship between disease risk and the optimal interval. For the validation, we use the MISCAN model as applied in a recently published CEA of actual screening policy guidelines in Switzerland [19].

METHODS

The mathematical models

We present two mathematical models of the cost-effectiveness of screening. The first is very simple and used only to introduce the modeling approach, while the second is more complete and employs more realistic assumptions. Both models are based on a natural history of cancer consisting of three phases (Figure 1) [20]. These phases are motivated by the possible detection of the tumor or its precursors by a screening test [21-23], with the example of breast cancer screening in mind.

Figure 1: Natural history of disease for 4 screened subjects, developing from no disease to a screen-detectable phase and on to the clinical phase. Screening occurs at t_i, screen detection at the diamond, clinical diagnosis at the circle and death at the arrow head. The lead time is shown between screen-detection and clinical diagnosis. Subject A is detected at t_3; B is detected at t_2 but dies from another cause and would have never entered the clinical phase; C is missed at t_2 due to imperfect test sensitivity, but detected at t_3; D is not screen-detected, due to a short preclinical phase.



The three stages in the model are (i) the "no disease" stage, in which a subject does not have the disease; (ii) the "screen-detectable" phase, during which the patient has preclinical disease that can be detected by the screening test; and, (iii) the "clinical" phase, by when the disease has been diagnosed due to symptoms. The benefits of a

screening program are realized by detecting and treating subjects in the screen-detectable phase, thereby postponing mortality in some subjects. To maintain a simple mathematical model we did not explicitly consider death from other causes during the natural history process.

The optimal screening interval depends on several parameters. Table 1 presents an overview of all parameters used in the mathematical models. The most important parameters can be described as follows.

1. The incidence rate (r) describes the annual probability that an individual will enter the preclinical phase. The incidence can depend on various risk factors, e.g. age at menarche, age at first live birth, number of previous biopsies, genetic predisposition and number of first-degree relatives with breast cancer [24]. For the mathematical models in this paper, we assume the incidence rate to be constant over age. The incident rate is assumed to be sufficiently low to satisfy the rare disease assumption.
2. The mean duration ($\overline{D}$) of the screen-detectable phase (in years, reflects the rate of disease progression).
3. The screening sensitivity (sn) represents the probability that a screening test will correctly identify an individual with preclinical disease as having cancer.
4. The screening specificity (sp) represents the probability that a screening test will correctly identify an individual without disease as disease-free.
5. The treatment effect (T) of a screen-detected case is the postponement of mortality due to early detection. The maximum treatment effect is achieved when a disease is detected as soon as it enters the preclinical phase. The mean treatment effect ($\overline{T}$) occurs when disease is screen-detected at the midpoint of the screen-detectable phase. $\overline{T}$ and T are expressed in quality adjusted life years (QALYs) gained.
6. The annual cost (C) of a screening program includes the fixed organizational costs of the program (c_0) and the variable costs (c) of a screening test. The fixed costs are sufficiently small that they can be disregarded. The variable cost, c, is given by

$$c = c_t + (1 - sp)c_d, \qquad (1)$$

where c_t is the cost of the test and c_d is the cost of diagnostic follow-up to a positive screen result. We make the simplifying assumptions that the cost of diagnostics only applies to false-positives and that the proportion of screening tests that are false-positive is 1-*sp*. We also assume that c_t, c_d and *sp* are constant, which implies that *c* is constant.

Table 1: Parameters used in the mathematical models

Parameter	Description
r	Annual incidence rate of preclinical cancer among patients in the absent phase
$\bar{D}$	The mean duration of the screen-detectable phase
D	The realized duration of the screen-detectable phase
sn	The screening sensitivity, i.e. the probability that a screen of an individual with preclinical disease leads to a cancer diagnosis
sp	The screening specificity, i.e. the probability that a screen of an individual without disease leads to a correct diagnosis
t	The time from the start of the screen-detectable phase
$\bar{T}$	The mean treatment effect, i.e. the number of QALYs gained when a cancer is screen-detected at the midpoint of the screen-detectable phase
T	The realized treatment effect, i.e. the number of QALYs gained for a specific screen-detected cancer
I	The interval (in years) between consecutive screenings
I_{opt}	The optimal interval (in years) between consecutive screenings, based on the cost-effectiveness threshold
p	Number of screenings during the screen-detectable phase
$E(D,I)$	Annual expected health effects (number of QALYs gained) as a function of the interval and the duration of the screen-detectable phase
C	Total costs of the screening program
c_0	Fixed cost of the screening program
c_t	Cost of screening test
c_d	Cost of diagnostics
c	Total variable costs per additional screening test
h	QALY decrement of a false positive screen result due to unnecessary diagnostics
H	Expected QALY decrement due to false-positive screen per test
$CERt$	Cost-effectiveness threshold, in euro per QALY gained

We first assume a constant treatment effect, a constant duration of the screen-detectable phase and perfect sensitivity and specificity; we refer to the resulting model as the simple model. Subsequently, we develop a second model that allows for a treatment effect that declines with the time spent in the screen-detectable phase, imperfect screening sensitivity and specificity and an exponential distribution of the duration of the screen-detectable phase; we call this model the detailed model.

Cost-effectiveness analysis

To determine the optimal screening interval I_{opt}, we used a cost-effectiveness (CE) perspective. CEA is the standard methodology for analyzing the costs and benefits of healthcare programs [25]. When screening programs of different intensities are compared, the incremental CE ratio (*ICER*) can be used to choose between the programs. The *ICER* is the ratio of the additional costs incurred (C) to the additional health effects gained (E) when a screening program is compared to less intensive screening programs or no screening at all:

$$ICER = \frac{\partial C}{\partial E} \qquad (2)$$

The *ICER* is typically compared to a CE threshold, which represents the marginal willingness to pay per QALY gained. We assume the existence of a threshold value given by *CERt*. A screening program is then acceptable if its *ICER* is not greater than the threshold *CERt*, i.e. $ICER \leq CERt$. Both costs and effects depend on I. Then from a CE perspective it must hold that

$$ICER = \frac{\partial C}{\partial E} = \frac{\partial C}{\partial I} \Big/ \frac{\partial E}{\partial I} \leq CERt\,. \qquad (3)$$

The optimal screening interval I_{opt}, is the interval with an *ICER* equal to the *CERt*.

The costs per individual per year, for a certain interval length, are the sum of fixed and interval dependent variable costs

$$C = c_0 + \frac{c}{I}. \qquad (4)$$

The derivative of equation 4 needed to calculate the *ICER* is

$$\frac{\partial C}{\partial I} = -\frac{c}{I^2}. \qquad (5)$$

This represents the change in costs per change in interval length. Equation 5 will be used throughout the paper.

Breast Cancer Example

The models are applied to breast cancer screening. Most of the parameter values in the model are specified to correspond to a published CEA of mammography screening in Switzerland [19]. We consider average-risk women between 50 and 70 years old, for whom the incidence rate for breast cancer is assumed 0.00225 per year, as assumed in the Swiss study [19]. We assume a mean duration of the screen-detectable phase of 3 years, in line with estimates for Canadian women over the age of 50 [26]. We assume a test sensitivity and specificity of 66% and 97% respectively, in line with estimates from a study of women aged between 50 and 69 used to inform the Swiss CEA [27]. The costs per screening test and costs of diagnostics following false-positive are assumed to be €138 and €650 respectively, also in line with the costs used in the Swiss CEA. The threshold, *CERt*, is set at 20,000 €/QALY [28]. While in reality the treatment effect per screen-detected cancer will depend on the screening interval, we assume a mean treatment effect of 4.2 QALYs per screen-detected cancer, approximating the QALY gain in the Swiss CEA. Finally, we also assume a QALY penalty of 0.001 for each false-positive, again in line with assumptions used in the Swiss CEA.

Verification

We verified our mathematical model against the MISCAN-Breast model. The MISCAN model was simplified significantly so that it matched the structural assumptions of the mathematical model and used the same parameter values. Using this simplified version of MISCAN we estimated the cost and effects of a range of screening programs with intervals ranging from 0.2 to 18 years over an 18 year screening program. From these estimates we identified the efficient frontier of strategies which represent the most effective strategies for a range of cost-effectiveness thresholds and the *ICER* of those efficient screening intervals.

The optimal screening interval was found with reference to the cost-effectiveness threshold. This was achieved by plotting the incremental cost-effectiveness ratio of the efficient screening intervals on the vertical axis and the corresponding length of those strategies on the horizontal axis. A horizontal line at a given value of the cost-effectiveness threshold was then imposed on the model. The interval length at the

intersection of the horizontal threshold line and the line connecting the ICERs of the efficient screening intervals. This process was repeated for a range of incidence rates from between one quarter to four times the incidence of the average risk group to derive the relationship between incidence and the optimal interval for comparison with the mathematical model.

Validation

We validated the detailed mathematical model against a full version of MISCAN-Breast used in a recently published CEA of organised mammography screening in Switzerland [19]. This full version of the MISCAN model did not use any of the structural simplifications used in the verification exercise. We assessed 22 different screening intervals ranging from 0.2 to 9 years in the fixed screening age range of 50 to 69 in previously unscreened women. The model is as described in de Gelder et al. [19], except that we simulated a single birth cohort rather than a population for the sake of simplicity.

We conducted a sensitivity analysis to investigate the effect of different cost-effectiveness thresholds on the model validation. We also conducted a sensitivity analysis to assess how the relationship between incidence rates and relative optimal screening interval varies with discount rates of 0%, 3% and 5% applied to costs and effects.

RESULTS

Model Results

Model 1: The Simple Model

The simple model assumes perfect test sensitivity and specificity (i.e. $sn = 1$ and $sp = 1$), a constant duration of the screen-detectable phase (so that $D = \bar{D}$) and constant incidence. The start of the screening program is assumed independent of the preclinical disease state among screening recipients. The effect of early detection of a cancer during the preclinical phase is considered constant (so that $T = \bar{T}$). The expected health effects per year for a screening interval I and a duration D of the screen-detectable phase are

$$\begin{aligned} E(D,I) &= \bar{T}r \quad \text{for} \quad I \le D, \\ E(D,I) &= \frac{\bar{T}rD}{I} \quad \text{for} \quad I > D. \end{aligned} \tag{6}$$

The *ICER* for the simple model is found by substituting the derivative of $E(D,I)$ over I and equation 5 into equation 3, which yields that

$$\begin{aligned} ICER &= \infty \quad \text{for} \quad I \le D \quad \text{where} \quad \frac{\partial E}{\partial I} = 0, \\ ICER &= \frac{c}{\bar{T}rD} \quad \text{for} \quad I > D \quad \text{where} \quad \frac{\partial E}{\partial I} = \frac{-\bar{T}rD}{I^2}. \end{aligned} \tag{7}$$

The *ICER* of ∞ for $I \le D$ reflects that all tumors will be found when $I = D$, and that further reduction of the interval therefore only adds costs, but no health effects. The *ICER* is constant for $I > D$, because c, r, D, and $\bar{T}$ are constants. The screening program will be socially acceptable if the *ICER* is less than or equal to the threshold (i.e. $CERt \ge c/\bar{T}rD$), in which case the optimal interval I equals the duration D of the screen-detectable phase.

Model 2: The Detailed Model

We now extend the simple model sequentially to consider: (i) a declining treatment effect of detection by screening with progression through the preclinical phase; (ii) imperfect screening sensitivity and specificity; (iii) an exponentially distributed duration of the screen-detectable phase. A summary of the assumptions of the simple and detailed models is shown in Table 2. We first show how the expected

health effects vary with the screening interval as each of these assumptions are added to the model. We then use the expression for the expected health effects with all three assumptions combined to estimate the *ICER* of different screening intervals and the change in the optimal screening interval as disease incidence varies.

Table 2: Summary of assumptions in the simple and detailed models

Model	Treatment Effect	Preclinical Duration	Screen Sensitivity	Screen Specificity
Simple	Constant	Constant	Perfect	Perfect
Detailed	Linearly Declining	Exponentially Distributed	Imperfect	Imperfect

Linearly Declining Screening Effect

We first assume the treatment effect of a screen-detected cancer declines linearly during the screen-detectable phase. This assumption is based on the fact that the probability that screen-detection prevents mortality for a cancer that would otherwise be lethal depends on the amount of lead time (the time between screen detection and the time when disease would have presented clinically in the absence of screening). If the lead time is 0, so that the time of screen-detection and the time of clinical detection coincide, no treatment effects should be expected. The duration of the screen-detectable phase is assumed to be constant and the screening sensitivity is assumed to be perfect. The treatment effect of screening at the beginning of the preclinical period is $2\overline{T}$ and declines linearly to 0 at the end of the period, leading to an average effect over the period of $\overline{T}$. Consequently the treatment effect for a screen-detected cancer is

$$T(t) = \frac{D-t}{D} 2\overline{T}, \qquad (8)$$

where t is the time between the start of the screen-detectable phase and the moment of screen-detection.

The expected annual health effects for a given interval and duration of the screen-detectable phase are as follows

$$E(D,I)=\frac{r}{I}\int_0^I T(t)dt=\frac{r}{I}\int_0^I \frac{D-t}{D}2\bar{T}dt=\frac{\bar{T}r(2D-I)}{D} \qquad \text{for } 0\le I\le D, \tag{9}$$

$$E(D,I)=\frac{r}{I}\int_0^D T(t)dt=\frac{r}{I}\int_0^D \frac{D-t}{D}2\bar{T}dt=\frac{\bar{T}rD}{I} \qquad \text{for } I>D.$$

Imperfect Screening Test Sensitivity and Specificity

We then add the assumption of imperfect screening sensitivity and specificity. The screening sensitivity and specificity are assumed constant during the preclinical phase. The probability that a preclinical cancer is detected at the i^{th} screening in the screen-detectable phase is

$$sn(1-sn)^{i-1}. \tag{10}$$

Imperfect specificity means there will be false-positive test results which will lead to a loss of utility. The expected utility decrement per screening is

$$H=(1-sp)h, \tag{11}$$

where the utility decrement of a false positive is given by h.

Combining equations 8 and 10 yields the expected treatment effects for the i^{th} screening during a screen-detectable phase

$$T(t_i)=\frac{D-t_i}{D}2\bar{T}sn(1-sn)^{i-1}, \tag{12}$$

where t_i is the time between the start of screen-detectable phase and the i^{th} screening. Because the interval between the screenings is constant, the i^{th} screening occurs at time

$$t_i=t_1+(i-1)I. \tag{13}$$

We first consider the case where $I \leq D$. In that case, D/I will have a value between two positive integers k and $k+1$, such that $kI \leq D < (k+1)I$, and the number of screens during the screen-detectable phase will be either k or $k+1$. Substituting equation 13 into 12 and subtracting the expected QALY decrement of a false-positive given by equation 11 shows the expected net health effect of the i^{th} screening to be

$$\frac{D-t_1-(i-1)I}{D} 2\bar{T}sn(1-sn)^{i-1} - H. \qquad (14)$$

Let p denote the number of screenings during the screen-detectable phase. The annual expected health effects, as a function of D and I can then be calculated as the expected treatment effect minus the expected cost QALY decrement of false-positives, which yields

$$E(D,I) = \frac{r}{I}\int_0^I \sum_{i=1}^{p} T(t_1)dt_1 - \frac{H}{I} =$$

$$\frac{2\bar{T}r}{I}\left[\int_0^{D-kI}\sum_{i=1}^{k+1}\frac{D-t_1-(i-1)I}{D}sn(1-sn)^{i-1}dt_1 + \int_{D-kI}^{I}\sum_{i=1}^{k}\frac{D-t_1-(i-1)I}{D}sn(1-sn)^{i-1}dt_1\right] - \frac{H}{I} =$$

$$\frac{2\bar{T}r}{I}\left[\int_0^{D-kI}\sum_{i=0}^{k}\frac{D-t_1-iI}{D}sn(1-sn)^{i}dt_1 + \int_{D-kI}^{I}\sum_{i=0}^{k-1}\frac{D-t_1-iI}{D}sn(1-sn)^{i}dt_1\right] - \frac{H}{I}, \quad \text{for } I \leq D \qquad (15)$$

where the first integral relates to when $p = k + 1$ screening intervals fall within the screen-detectable phase and the second integral relates to when $p=k$ screenings fall within the screen-detectable phase. The derivation of equation 16 from 15 uses the approximation $k \approx D/I$ and is shown in Appendix I.

$$E(D,I) = \frac{Tr}{I}\left[\frac{I^2}{snD}\left(1-(1-sn)^{\frac{D}{I}}\right)(sn-2)+2I\right] - \frac{H}{I} \qquad \text{for } I \leq D. \qquad (16)$$

In the case in which $I > D$ there is at most 1 screening (at time t_1) during the screen-detectable phase, and the expected health effects can be calculated as

$$E(D,I) = \frac{r}{I}\int_0^D E(t_1)\,dt_1 - \frac{H}{I}$$

$$= 2\bar{T}\frac{r}{I}\int_0^D \frac{D-t_1}{D}sn\,dt_1 - \frac{H}{I} = \frac{\bar{T}rsnD - H}{I} \qquad \text{for } I > D. \qquad (17)$$

Exponentially Distributed Preclinical Phase

We now assume that the duration of the preclinical phase is exponentially distributed, so that the probability density function is $f(D) = \frac{1}{\bar{D}} e^{-D/\bar{D}}$. The annual expected health effects of screening for $I \leq D$ and $I > D$ from equations 16 and 17 are now combined with the exponential distribution of the screen-detectable phase, which shows that the annual expected health effects as a function of the screening interval to be

$$E(I) = \int_0^\infty f(D)E(D,I)dD =$$

$$\int_0^I \frac{1}{\bar{D}} e^{-D/\bar{D}} \frac{\bar{T}rsnD - H}{I} dD +$$

$$\int_I^\infty \frac{1}{\bar{D}} e^{-D/\bar{D}} \left[\frac{\bar{T}r}{I} \left[\frac{I^2}{snD} \left(1 - (1 - sn)^{\frac{D}{I}} \right)(sn - 2) + 2I \right] - \frac{H}{I} \right] dD =$$

$$\frac{\bar{T}r}{I} \left[\int_0^I snD \frac{1}{\bar{D}} e^{-D/\bar{D}} dD + \int_I^\infty \frac{I^2 \left(1 - (1 - sn)^{\frac{D}{I}} \right)(sn - 2) + 2snDI}{snD} \frac{1}{\bar{D}} e^{-D/\bar{D}} dD \right] - \frac{H}{I}, \quad (18)$$

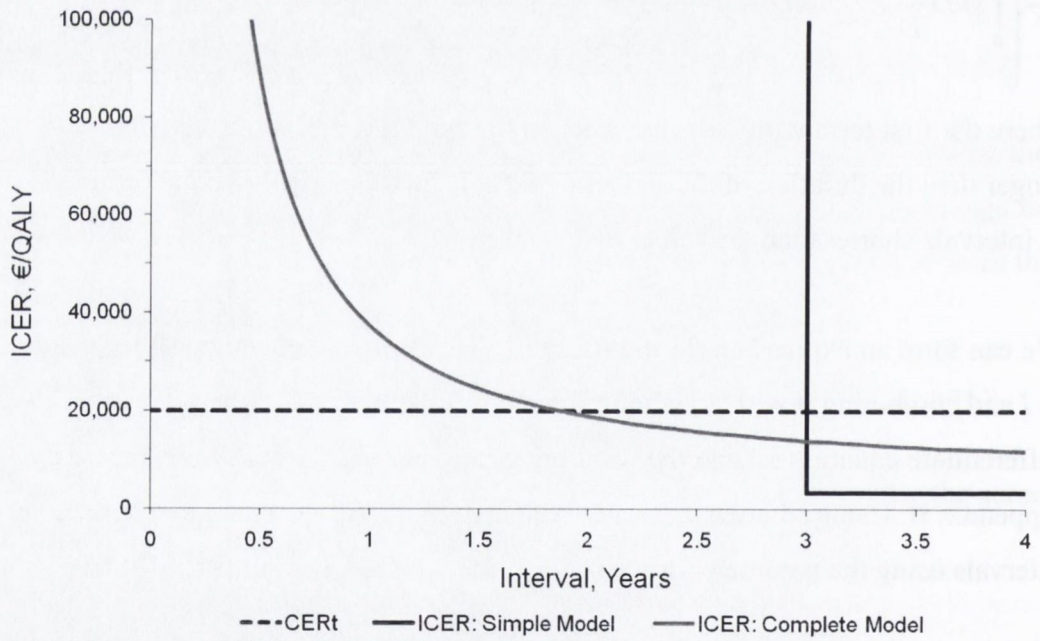
where the first term within the brackets on the final line refers to screening intervals longer than the duration of the screen-detectable phase, and the second term refers to intervals shorter than the screen-detectable phase.

We can form an expression for the *ICER* by differentiating equation 18 with respect to I and combining it with equations 3 and 5. Mathematica 4.0 was used to differentiate equation 18 and the resulting expression for the *ICER* is reported in Appendix II. Using Mathematica, we evaluated the *ICER* for a range of screening intervals using the parameter described above.

Example

First consider the simple model. Imputing the assumed values for breast cancer screening in equation 7 results in the relationship between the *ICER* and the interval length shown with the black line in Figure 2. The *ICER* intersects *CERt* at the value *D* and therefore I_{opt} is 3 years. Now considering the detailed model, the derived function for the *ICER* was evaluated over a range of interval lengths, resulting in the grey line in Figure 2. In this case the intersection of the *ICER* curve and the *CERt* indicates an I_{opt} of approximately 1.8 years. The incidence rate was then varied to generate new *ICER* curves, which were combined with a threshold of €20,000 per QALY, to determine the relationship between incidence, *r,* and I_{opt} shown with the grey line in Figure 3B.

Figure 2: Variation of the *ICER* with the screening interval under the simple and the detailed model with a cost-effectiveness threshold of €20,000/QALY



Verification

The results of the verification analysis are shown graphically in Figures 3A and B. The black diamonds in Figure 3A show the *ICER* estimates at different incidence rates from the simplified MISCAN model, while the grey line shows the *ICER* estimates from the mathematical model. The figure shows the *ICERs* from the MISCAN and analytic models to be very close. The *ICERs* from the MISCAN model are consistently marginally lower than those from the mathematical model. This is to be expected, as the MISCAN model estimates *ICERs* based on discrete changes to the screening interval, while the mathematical model considers marginal changes, which necessarily results in higher *ICERs*. The relationship between the incidence rate and the optimal interval is shown in Figure 3B. The black diamonds and grey line show the optimal interval estimates at different incidence rates from the simplified MISCAN and mathematical models respectively. Again the results indicate a close match between the two models. As before, the small remaining difference between the estimates is explained by the fact that the MISCAN model simulates discrete changes to the interval, while the interval changes continuously in the mathematical model.

Validation

The optimal screening intervals estimated by the mathematical model did not exactly match those derived from the complete MISCAN model. However, the purpose of our mathematical model is to show the relative change in optimal screening intensity with a change in disease incidence, rather than to estimate the optimal level of screening intensity at any one point. Consequently, we expressed the optimal screening interval at each incidence rate as an index relative to the optimal interval at the average-risk incidence rate. The relationship between the incidence rate and the relative optimal screening interval is shown in Figure 4, with the grey line for the mathematical model and with the black diamonds for the MISCAN model. It shows a close agreement between the models on the relationship between the incidence rate and the relative screening intensity, although there is some deviation at incidence rates below the baseline average-risk.

Figure 3A: Verification of the detailed model against a simplified version of the MISCAN model showing the variation of the *ICER* with the screening interval

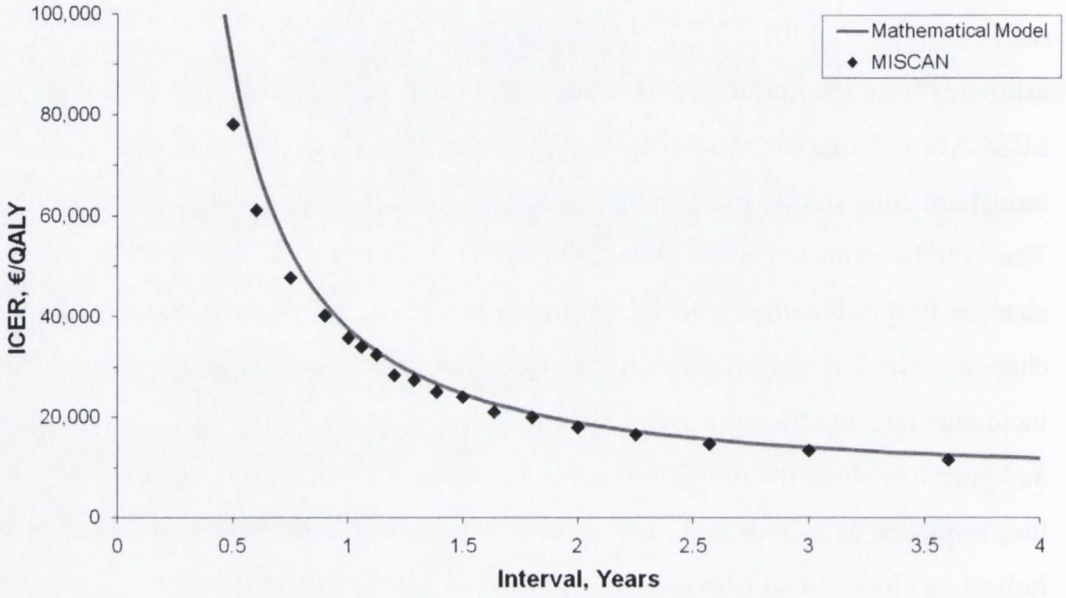
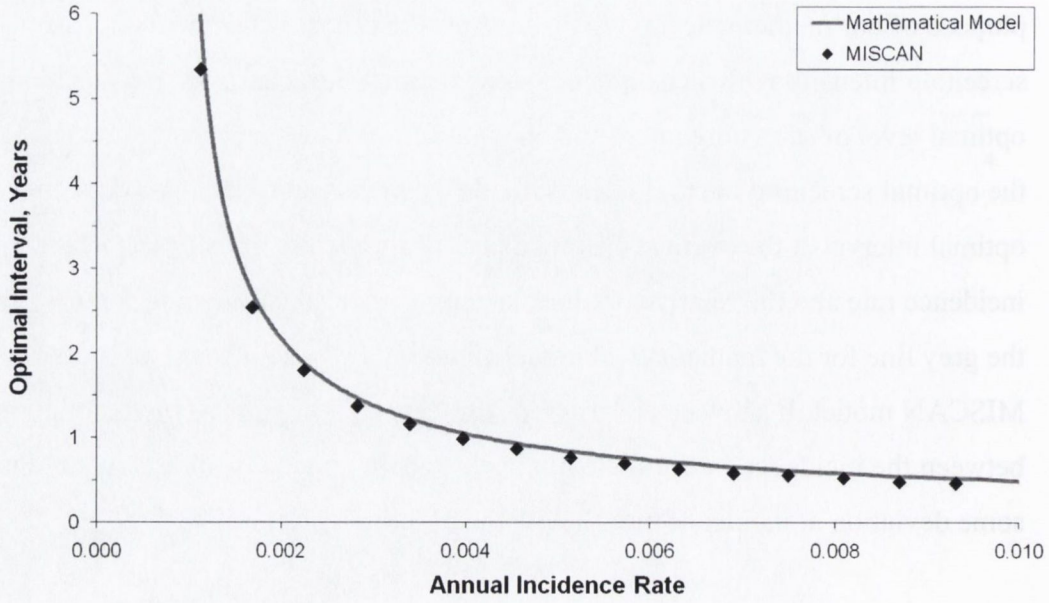
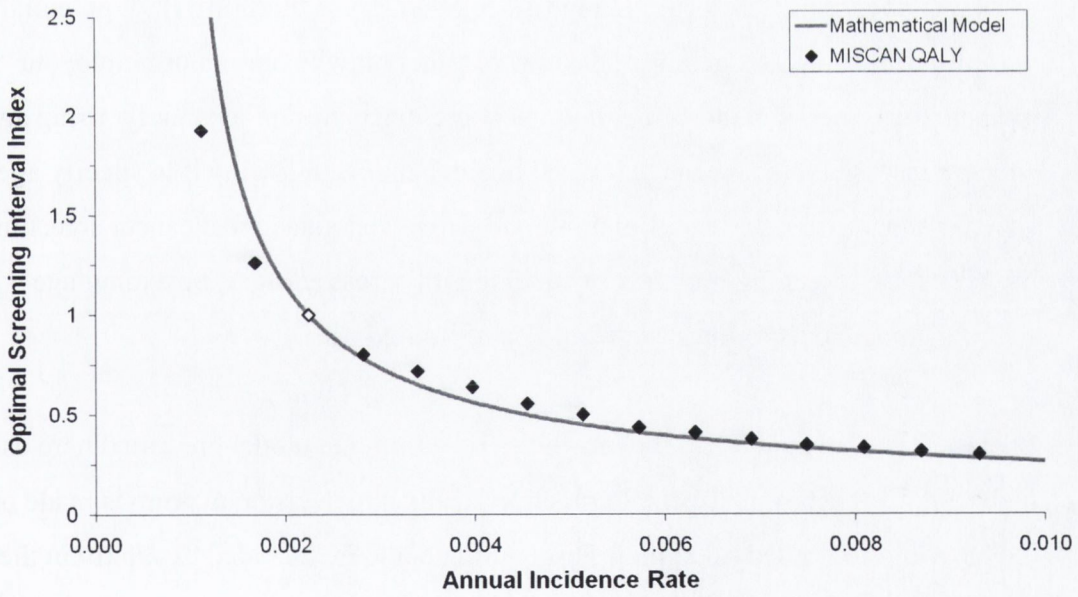


Figure 3B: Verification of the detailed model against a simplified version of the MISCAN model showing the variation of the optimal screening interval at a cost-effectiveness threshold of €20,000/QALYwith the annual incidence rate



The results of a sensitivity analysis for different levels of the cost-effectiveness threshold are shown in Figure 5 in Appendix III. The relative change in the optimal interval with changing incidence is shown for thresholds of 20, 30 and 50 thousand €/QALY, each relative to the case at the 20,000 €/QALY threshold. The results show the match was maintained at these other threshold levels.

Figure 4: Validation of the detailed model against a complete version of the MISCAN model showing the relative change in the optimal screening interval at a cost-effectiveness threshold of €20,000/QALY with the annual incidence rate and with the optimal screening interval indexed to 1 at the baseline annual incidence rate of 0.00225 and sensitivity analysis showing the relative change in the optimal screening interval with the incidence rate under the assumption of perfect test specificity



DISCUSSION

Using a mathematical model we estimated the relationship between the incidence rate and the length of the optimal screening interval. For the simple model, the optimal interval is equal to the preclinical duration of the disease of interest. For the detailed model, the optimal interval became shorter. Our results show that a mathematical model can provide a clear insight into how disease risk influences the optimal screening interval. Furthermore, it shows how a small number of parameters can be used to estimate the change in the optimal screening interval for individual risk profiles.

Reality is more complex than represented in our detailed model. For example, the incidence of breast cancer increases with age, while our mathematical model assumes it to be constant. This would suggest the optimal screening interval would become shorter with age. However, the potential benefit of early detection decreases with age, due to increased competing risk of other-cause mortality [29], potentially partially offsetting a reduction in the optimal interval with age. In principle, our mathematical model could be applied to different age groups separately, using age-specific disease incidence and expected health benefits, allowing it to specify age-specific adjustments to the optimal interval. However, since most cancer screening programs use screening intervals of equal length across all ages, providing age-specific adjustments to intervals would be of limited use.

Like the other mathematical models in the literature, the model presented here uses undiscounted costs and effects. In practice, healthcare decision making is made on the basis of discounted outcomes. Discounting could be included explicitly in the model, but at the cost of additional complexity. We conducted a sensitivity analysis to assess how the optimal screening interval varies with disease incidence using the MISCAN model when discount rates of 3 % and 5% are used for both costs and effects (results not shown). We found that although the length of the optimal interval changed with discounting, the relative relationship between incidence and the optimal interval index did not vary substantially.

The advantage of the mathematical model is that it is a fast and easy to use method for understanding the influence of relative disease risk on the optimal screening intervals. It requires few input parameters and yields estimates quickly. As such, it offers a useful means for interpreting the impact of biomarkers as potential risk predictors on the optimal interval [30]. In particular, it offers a simple way of assessing the significance of a large numbers of biomarkers with a cumulative effect on the optimal screening interval, without the need for separate analyses of each biomarker separately.

It should be recognized that there are two substantial differences between the complete MISCAN model and the mathematic model compared in this analysis. The mathematical model is solved analytically rather than using microsimulation and it is a much simpler model, employing less parameters and having a far simpler structure. What has been achieved here with the simpler mathematical model could probably have also been done with a simplified microsimulation model. However, such a simplified microsimulation model would not have the benefit of having a continuous relationship between incidence and the optimal screening interval that can be solved analytically.

CONCLUSION

We have shown that a relatively simple mathematical model can be used to determine the optimal screening interval for high-risk groups relative to the optimal screening interval for the average-risk population. Despite the simplifying assumptions of the mathematical model, the validation exercise shows that it is sufficient to capture the effect of changes in disease incidence on optimal screening intervals. Consequently, the mathematical model presented here offers a quick and easy way to explore optimal screening intervals for specific risk groups without the use of complex simulation models.

APPENDIX I

Derivation of equation 16 from 15.

$$E(D,I)=\frac{r}{I}\int_0^I\sum_{i=1}^{n}E(t_1)dt_1-\frac{H}{I}=$$

$$\frac{2\bar{T}r}{I}\left[\int_0^{D-kI}\sum_{i=1}^{k+1}\frac{D-t_1-(i-1)I}{D}sn(1-sn)^{i-1}dt_1+\int_{D-kI}^{I}\sum_{i=1}^{k}\frac{D-t_1-(i-1)I}{D}sn(1-sn)^{i-1}dt_1\right]-\frac{H}{I}$$

Subtracting i by one to simplify the summations in both integrals.

$$\frac{2\bar{T}r}{I}\left[\int_0^{D-kI}\sum_{i=0}^{k}\frac{D-t_1-iI}{D}sn(1-sn)^{i}dt_1+\int_{D-kI}^{I}\sum_{i=0}^{k-1}\frac{D-t_1-iI}{D}sn(1-sn)^{i}dt_1\right]-\frac{H}{I}$$

Applying the approximation $k\approx D/I$ results in $D\text{-}kI\approx 0$, allowing the elimination of the first integral and the lower bound of integration on the second becomes zero. Expanding the three terms of the ratio within the second integral in the following.

$$\frac{2\bar{T}r}{I}\left[\int_0^{I}\sum_{i=0}^{k-1}sn(1-sn)^{i}dt_1+\int_0^{I}\sum_{i=0}^{k-1}\frac{-t_1}{D}sn(1-sn)^{i}dt_1+\int_0^{I}\sum_{i=0}^{k-1}\frac{-iI}{D}sn(1-sn)^{i}dt_1\right]-\frac{H}{I}$$

Expanding the summations of the first two integrals.

$$\frac{2\bar{T}r}{I}\left[\int_0^{I}\left(1-(1-sn)^{k}\right)dt_1+\int_0^{I}\frac{-t_1}{D}\left(1-(1-sn)^{k}\right)dt_1+\int_0^{I}\sum_{i=0}^{k-1}\frac{-iI}{D}sn(1-sn)^{i}dt_1\right]-\frac{H}{I}$$

Integrating all three integrals, gathering terms from the first two and bringing 2 within the square brackets.

$$\frac{\bar{T}r}{I}\left[\left(2I-\frac{I^2}{D}\right)\left(1-(1-sn)^{k}\right)+2\sum_{i=0}^{k-1}\frac{-iI^2}{D}sn(1-sn)^{i}\right]-\frac{H}{I}$$

Expanding the summation and factoring terms.

$$E(D,I)=\frac{\bar{T}r}{I}\left[\left(2I-\frac{I^2}{D}\right)\left(1-(1-sn)^k\right)-\frac{2snI^2}{D(1-sn)}\Big/\left(\frac{1}{1-sn}-1\right)\left(\left(\frac{1}{(1-sn)^k}-1\right)\Big/\left(\left(\frac{1}{1-sn}-1\right)\frac{1}{(1-sn)^k}\right)-k(1-sn)^k\right)\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[\left(2I-\frac{I^2}{D}\right)\left(1-(1-sn)^k\right)-2\frac{I^2}{D}\left(\left(\frac{1}{(1-sn)^k}-1\right)\Big/\left(\frac{sn}{1-sn}\frac{1}{(1-sn)^k}\right)-k(1-sn)^k\right)\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[\left(2I-\frac{I^2}{D}\right)\left(1-(1-sn)^k\right)-2\frac{I^2}{D}\left(\frac{(1-sn)\left(1-(1-sn)^k\right)}{sn}-k(1-sn)^k\right)\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[\left(2I-\frac{I^2}{D}\frac{sn+2(1-sn)}{sn}\right)\left(1-(1-sn)^k\right)+2\frac{I^2}{D}k(1-sn)^k\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[\left(2I+\frac{I^2}{D}\frac{sn-2}{sn}\right)\left(1-(1-sn)^k\right)+2I(1-sn)^k\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[2I+\frac{I^2}{D}\frac{sn-2}{sn}\left(1-(1-sn)^k\right)\right]-\frac{H}{I}=$$

$$\frac{\bar{T}r}{I}\left[\frac{I^2}{snD}\left(1-(1-sn)^{\frac{D}{I}}\right)(sn-2)+2I\right]-\frac{H}{I}$$

APPENDIX II

An expression for the *ICER* in the complete model is formed by differentiating equation 18 with respect to I and combining with equations 3 and 5. Eliminating the indefinite terms that tend towards zero as the upper limit of integration on the second term within the brackets of equation 18 tends to infinity results in the following expression for the *ICER*

$$\mathrm{ICER} = -\frac{c}{\overline{\overline{T}}r}\left[\begin{array}{l}\dfrac{H}{\overline{\overline{T}}r} + sn\left(\mathrm{e}^{-I/\overline{D}}\left(\overline{D}+I\right)-\overline{D}\right)+ \\ \dfrac{I^2(-2+sn)}{sn\overline{D}}\left(\mathrm{Ei}\left(-\dfrac{I}{\overline{\overline{D}}}+\ln(1-sn)\right)-\mathrm{Ei}\left(-\dfrac{I}{\overline{\overline{D}}}\right)+\mathrm{e}^{-I/\overline{D}}(sn-1)\left(1+\dfrac{I}{\overline{D}\ln(1-sn)-I}\right)\right)\end{array}\right]^{-1}$$

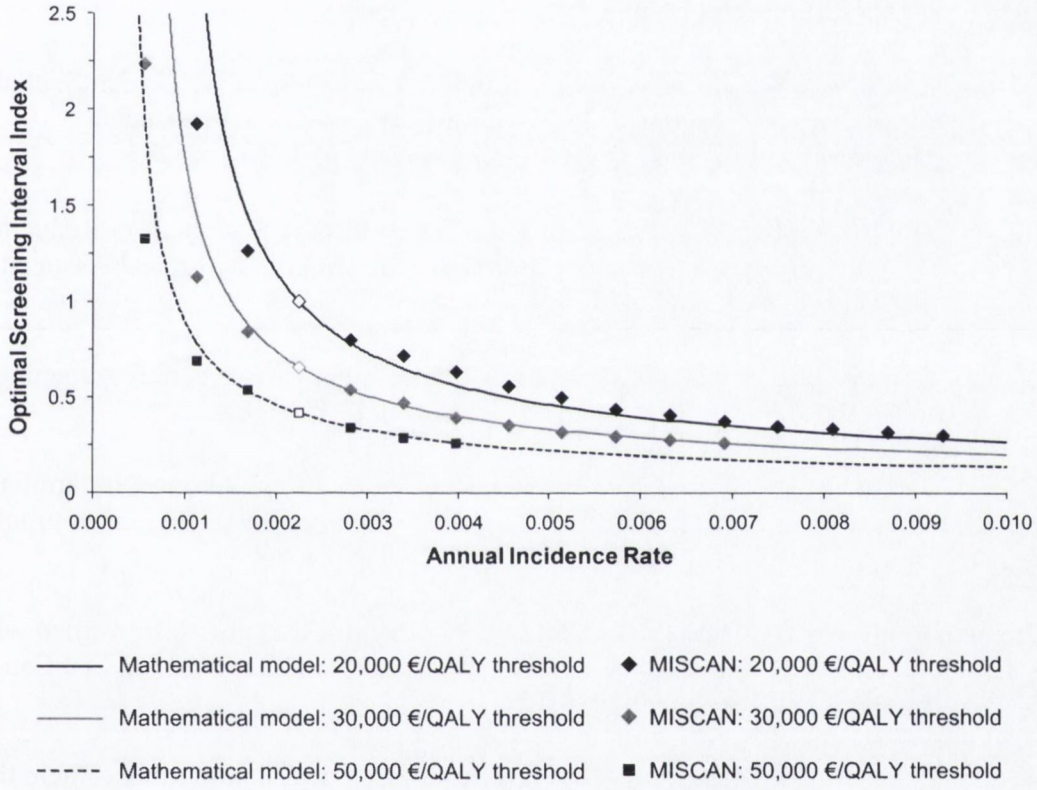
where the function Ei() denotes the exponential integral, which is defined as

$$\mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{\mathrm{e}^t}{t}dt\,.$$

An expression for the optimal interval I_{opt} in terms of the threshold *CERt* cannot be found. However, the incidence rate r can be expressed in terms of the I_{opt} at the cost-effectiveness threshold, so the incidence-optimal interval relationship is described in this way.

APPENDIX III

Figure 5: Sensitivity analysis of the relative change in the optimal screening interval with disease incidence at thresholds of 20, 30 and 50 thousand €/QALY in the detailed mathematical and the complete MISCAN models, each relative to the base case threshold of 20,000 €/QALY



REFERENCES

1. Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, et al. The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. British Journal of Cancer 1994;70(3):542-8.

2. Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. British Journal of Medicine 1988;297(6654):943-8.

3. Nyström L, Wall S, Rutqvist LE, Lindgren A, Lindqvist M, Rydén S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. The Lancet 1993;341(8851):973-8.

4. Shapiro S, Strax P, Venet L. Periodic Breast Cancer Screening in Reducing Mortality From Breast Cancer. Journal of the American Medical Association 1971;215(11):1777-85.

5. Baker R. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Management Science 1998;1(2):103-13.

6. Brandt A, Bermejo JL, Sundquist J, Hemminki K. Age of onset in familial breast cancer as background data for medical surveillance. British Journal of Cancer 2009;102(1):42-7.

7. Brandt A LBJ, Sundquist J, Hemminki K. Breast cancer risk in women who fulfill high-risk criteria: at what age should surveillance start? Breast Cancer Research and Treatment 2010;121(1):133-41.

8. Knudsen AB, McMahon PM, Gazelle GS. Use of Modeling to Evaluate the Cost-Effectiveness of Cancer Screening Programs. Journal of Clinical Oncology 2007;25(2):203-8.

9. Ramsey SD, Wilschut J, Boer R, van Ballegooijen M. A Decision-Analytic Evaluation of the Cost-Effectiveness of Family History-Based Colorectal Cancer Screening Programs. American Journal of Gastroenterology 2010;105(8):1861-9.

10. Habbema JDF, Lubbe JTN, van Oortmarssen GJ, van der Maas PJ. A simulation approach to cost-effectiveness and cost-benefit calculations of screening for the early detection of disease. European Journal of Operational Research 1987;29(2):159-66.

11. Habbema JDF, van Oortmarssen GJ, Lubbe JTN, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine 1985;20(1):79-93.

12. Van Oortmarssen GJ, Habbema JDF, Van Der Maas PJ, De Koning HJ, Collette HJA, Verbeek ALM, et al. A model for breast cancer screening. Cancer 1990;66(7):1601-12.

13. Kirch RLA, Klein M. Examination schedules for breast cancer. Cancer 1974;33(5):1444-50.

14. Zelen M. Optimal scheduling of examinations for the early detection of disease. Biometrika 1993;80(2):279-93.

15. Lee SJ, Zelen M. Scheduling Periodic Examinations for the Early Detection of Disease: Applications to Breast Cancer. Journal of the American Statistical Association 1998;93(444):1271-81.

16. Parmigiani G. On Optimal Screening Ages. Journal of the American Statistical Association 1993;88(422):622-8.

17. Parmigiani G. Timing medical examinations via intensity functions. Biometrika 1997;84(4):803-16.

18. Parmigiani G, Kamlet M. A Cost-Utility Analysis Of Alternative Strategies In Screening For Breast Cancer. In: Bayesian Statistics in Science and Technology: Case Studies, 1993.

19. de Gelder R, Bulliard J-L, de Wolf C, Fracheboud J, Draisma G, Schopper D, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. European Journal of Cancer 2009;45(1):127-38.

20. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. America Journal of Epidemiology 1983;118(6):865-86.

21. Gao F, Chia K-S, Ng F-C, Ng E-H, Machin D. Interval cancers following breast cancer screening in Singaporean women. International Journal of Cancer 2002;101(5):475-9.

22. Jansen JTM, Zoetelief J. MBS: a model for risk benefit analysis of breast cancer screening. British Journal of Radiology 1995;68(806):141-9.

23. Jansen JTM, Zoetelief J. Optimisation of mammographic breast cancer screening using a computer simulation model. European Journal of Radiology 1997;24(2):137-44.

24. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. Journal of the National Cancer Institute 1989;81(24):1879-86.

25. Weinstein MC, Stason WB. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. New England Journal of Medicine 1977;296(13):716-21.

26. Shen Y, Zelen M. Screening Sensitivity and Sojourn Time From Breast Cancer Early Detection Clinical Trials: Mammograms and Physical Examinations. Journal of Clinical Oncology 2001;19(15):3490-9.

27. Bulliard JL, De Landtsheer JP, Levi F. Results from the Swiss mammography screening pilot programme. European Journal of Cancer 2003;39(12):1761-9.

28. Boersma C, Broere A, Postma MJ. Quantification of the Potential Impact of Cost-effectiveness Thresholds on Dutch Drug Expenditures Using Retrospective Analysis. Value in Health 2010;13(6):853-6.

29. Brown ML. Economic Considerations in Breast Cancer Screening of Older Women. The Journals of Gerontology 1992;47(Special Issue):51-8.

30. D'Souza AL, Tseng JR, Pauly KB, Guccione S, Rosenberg J, Gambhir SS, et al. A strategy for blood biomarker amplification and localization using ultrasound. Proceedings of the National Academy of Sciences of the United States of America 2009;106(40):17152-7.

Paper 6.

The Cost-Effectiveness of Human Papillomavirus Vaccination: The Relevance of Alternative Screening Intensities

James F. O'Mahony, MA[1,2],
Inge M. C. M. de Kok, MSc[2],
Joost van Rosmalen, PhD[2],
Marjolein van Ballegooijen, MD, PhD[2].

1. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland;
2. Department of Public Health, Erasmus University Medical Centre, Rotterdam, the Netherlands;

ABSTRACT

Background: Many cost-effectiveness analyses of vaccination against the Human Papillomavirus do not consider cervical screening strategies other than the current programme. This is despite awareness in the literature that the cost-effectiveness of vaccination depends in part on screening intensity and that all relevant alternative interventions should be considered, including alternative screening intensities. Consequently, CEAs that only consider the current screening intensity can overlook the effect of alternative screening strategies on the vaccine's cost-effectiveness and threshold price.

Objective: To assess the consequences of considering alternative screening intervals and start and stop ages in a CEA model previously used to appraise HPV vaccination in the Netherlands.

Methods: A review of published CEAs of HPV vaccination was conducted to establish how many existing analyses considered alternative screening intensities. A microsimulation model used in a previously published CEA of HPV vaccination was adapted to consider alternative screening intensities.

Results: Many but not all published CEAs of HPV vaccination only consider the current screening intensity. Those that considered alternative screening intensities tended to be from either low or middle income countries or the US. Assuming Dutch cervical screening remains at its current intensity yields a cost-effectiveness ratio of adding vaccination to screening of €28,500 per life year gained (LYG), while considering a large range of alternative screening intensities results in a higher ratio of €35,400/LYG.

Conclusion: Considering alternative screening intensities is relevant to CEAs of HPV vaccination in general and led to a less favourable cost-effectiveness estimate in the Dutch context. The consideration of alternative screening intensities prompts a reconsideration of how the cost-effectiveness of vaccination should be interpreted from the given cost and effects estimates.

INTRODUCTION

Three CEAs of vaccination against HPV for the prevention of cervical cancer in the Netherlands were published in 2009 [1-3]. All three compared vaccination in addition to the current Dutch cervical screening strategy of 7 lifetime screens using pap cytology at 5 year intervals from age 30 to 65 to screening alone. This assumption that screening remains unchanged with the introduction of vaccination is common to many CEAs of HPV vaccination [1]. This is despite an awareness in the literature that the cost-effectiveness (CE) of vaccination depends in part on the intensity of screening [2]. It is also widely recognised that CEAs in general should consider all relevant alternatives [3]. It has been recognised that, in the particular context of HPV vaccination, the relevant alternatives include alternative screening intensities [4].

The purpose of this analysis is to illustrate the consequences of considering alternative screening intervals and start and stop ages in one of the three Dutch CEAs of HPV vaccination. We show how the consideration of multiple screening alternatives requires a reinterpretation of the incremental cost-effectiveness of HPV vaccination since the vaccine's cost-effectiveness is partly contingent on the intensity of screening [5]. In addition, we provide context to our analysis by reviewing the comparisons made in other published CEAs of HPV vaccination in the Netherlands and elsewhere.

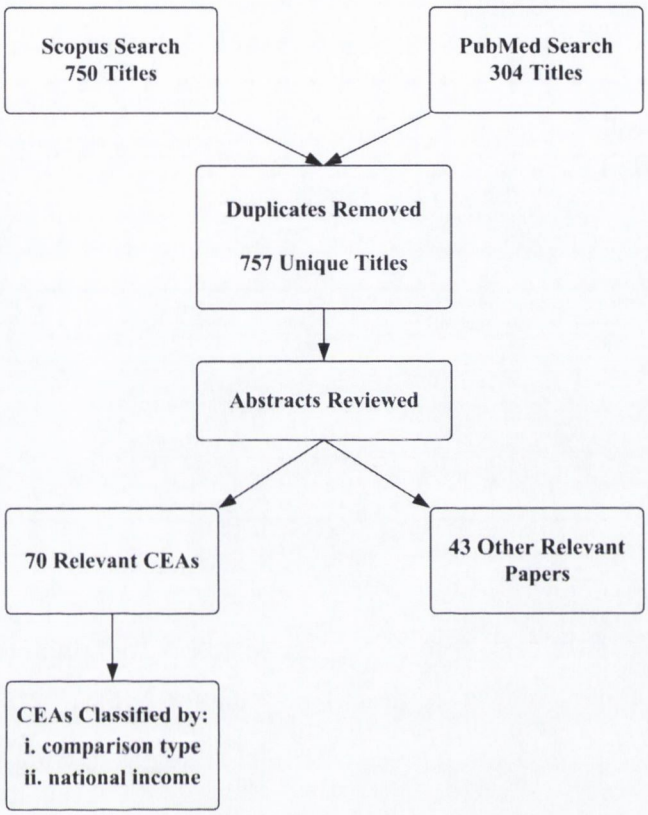
This study is organised as follows. The literature review section describes the search criteria used in the review and reports the findings. The methods section describes the model used to estimate the vaccine's cost-effectiveness with and without multiple alternative screening strategy comparators. The results section reports the findings of the review and the cost-effectiveness estimates in a CE-plane. The discussion section interprets the result of the literature review and the CEA, including a consideration of what the inclusion of alternative screening strategies implies for our understanding of what is the ICER of the HPV vaccine.

LITERATURE REVIEW

A review of the literature was performed through the Elsevier's SciVerse Scopus database (www.scopus.com) and the US National Library of Medicine's PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) to find CEAs of HPV vaccination for the prevention of cervical cancer. The databases were searched for titles, abstracts and keywords matching the following string: (HPV OR "human papilloma*" OR "human papilloma*" OR cervical) AND (vaccin* OR immun*) AND ("cost effect*" OR "cost-effect*" OR "cost utility" OR "cost-utility" OR "health technology assessment" OR CEA OR CUA OR HTA). The search was restricted to journal articles published in English from 2000 and onwards. The search was conducted on July 9th, 2012.

Figure 1. summarises the search process. The search results from the two databases were pooled and duplicates removed. The resulting unique abstracts were then hand sorted by one reviewer to identify relevant studies. Of most interest were those studies that were CEAs of HPV vaccination. The review also considered other studies that were not explicitly CEAs but contained relevant material, such as reviews and commentaries. All the studies considered relevant on the basis of their abstracts were then reviewed, with the CEAs further classified as either (i) those that estimated vaccine cost-effectiveness assuming screening remained unchanged; (ii) those that considered alternative screening intensities among the vaccination-plus-screening scenarios only; (iii) those that considered alternative screening intensities in both the screening-only and vaccination-plus-screening scenarios. The studies were also classified as either high income or middle and low income countries according to the International Monetary Fund's (IMF) classification of states by income [6].

Figure 1. The Literature Review Process



The review yielded 757 unique titles, of which 70 were identified as relevant CEAs. Of the 70 CEAs, 46 estimated the cost-effectiveness of vaccination as the incremental cost-effectiveness of adding vaccination to current screening [5, 7-50]. There were 3 studies that accounted for a range of alternative screening intensities among the vaccination-plus-screening strategies, but did not consider alternative screening-only strategies [51-53]. There were 17 CEAs that considered alternative screening strategies in both the screening-only and vaccination-plus-screening options, estimating cost-effectiveness as the frontier over all efficient strategies [4, 54-69]. One of the 17 studies is notable for reporting both the ICER of adding vaccination to current screening and reporting estimates for an efficient frontier made up of alternative screening and vaccination strategies [4]. Four studies made alternative interpretations of their estimates that do not fit the classification used here [70-73]. These findings are summarized in Table 1.

Table 1. Range of alternative screening strategies compared when estimating the cost-effectiveness of adding HPV vaccination to cervical cancer control programmes decomposed by IMF income category

IMF income category	Range of alternative screening strategies considered			
	None	Screening plus vaccination only	Full	Other comparison
High	38	3	6	3
Middle or low	8	0	11	1
Total	**46**	**3**	**17**	**4**

There was an evident distinction between the studies' countries of origin given the alternatives considered. Of the 46 that assumed screening remained unchanged, 38 were from high income countries, of which 27 were European. Conversely, of the 17 studies that considered alternative screening strategies, 6 were from high income countries, only 2 of which were European.

METHODS

This study employs the same CEA model of cervical cancer prevention used by de Kok et al. in their appraisal of HPV vaccination [7]. The model used is the MISCAN microsimulation model of cervical cancer prevention developed at Erasmus MC, the Netherlands [74]. The model simulates the individual life histories of a birth cohort of women. In the absence of vaccination and screening the simulated women develop cervical lesions at a certain rate, which can progress to cancer, potentially leading to cervical cancer death. Treatment costs are applied to the different interventions and health states in the model, permitting estimation of the interventions' net costs and health effects. The health effects of cervical cancer prevention are estimated here in terms of life years gained. Further detail of the model description and its assumptions can be found in de Kok et al. [7].

The disease incidence rates and sojourn times employed in the model are derived from an age-period-cohort model of cervical cancer in the years before the introduction of screening in the Netherlands [75]. The model assumptions for the age-period-cohort analysis are supplemented by calibrating the model to data on the observed rates of cervical intraepithelial neoplasia (CIN) within the screened population between 1997 and 2001 since the introduction of screening [76] and the observed rates of age-specific infections with high risk HPV types [77].

Adding vaccination against HPV types 16 and 18 reduces the simulated incidence of disease incidence, while adding screening reduces disease progression by facilitating early treatment of screen-detected lesions. The vaccine is assumed to be 70% effective against cancer and 35% effective against pre-invasive lesions. These assumptions are based on the prevalence of the high risk HPV types 16 and 18 cancers and pre-invasive lesions coupled with the assumption of lifelong vaccine protection [77]. The early detection of screening was modelled by applying the assumed test sensitivities of 50%, 65%, and 80% for CIN grades 1 to 3 respectively. Test sensitivity was assumed to be 80% and 85% respectively for preclinical invasive cancer stages I and II+ respectively [7]. A specificity rate of 98.5% was based on the observed false positive rate in the Dutch screening programme [7].

The original analysis by de Kok et al. assumed screening as per the current Dutch screening programme when estimating vaccine cost-effectiveness. We adapt the original CEA to simulate multiple alternative screening strategies, following van den Akker-van Marle et al.'s approach to the simulation of screening alternatives in MISCAN [78]. We simulate over 5,000 alternative screening strategies with screening start and stop ages varying between 17 and 70 with the number of lifetime screens varying between 1 and 54. All screening intervals are assumed constant throughout the screening programme and all screens used conventional pap cytology as both the primary and triage screen tests. Our analysis differs slightly from the original analysis regarding screening coverage assumptions, as the original analysis applied observed screening rates to the current programme, whereas this analysis applied a common assumption of 80% screening coverage and 80% screening follow up. This simplifying assumption was required to provide a common basis for comparison between alternative screening strategies.

The model used here follows de Kok et al.'s assumption of an 85% vaccine coverage rate, with no assumed correlation to disease risk or screening behaviour. The vaccine's cost per dose is the same as assumed in the original analysis of €118. We also consider a higher and lower vaccine price of €200 and €50 a dose. The vaccine is administered to girls aged 12, in accordance with the current Dutch vaccination programme and the original analysis. Costs and health effects are discounted to the year of vaccination. The discount rates used are those recommended by Dutch CEA guidelines of 4% and 1.5% for costs and health effects respectively [79]. A birth cohort of 1 million women is used in the model, while the reported results are rescaled to a cohort of 100,000 women, which approximates the annual number of 12 year old girls eligible for vaccination in the Netherlands.

The model's cost and effects estimates are presented in a cost-effectiveness plane. The first comparison of cost-effectiveness is that made by de Kok et al. of the incremental cost-effectiveness ratio of adding vaccination to current screening. The second comparison is between all the simulated strategies, both the screening-only strategies and the vaccination-plus-screening strategies. The second comparison is used to identify the efficient frontier of interventions and find the ICERs between each of the efficient interventions.

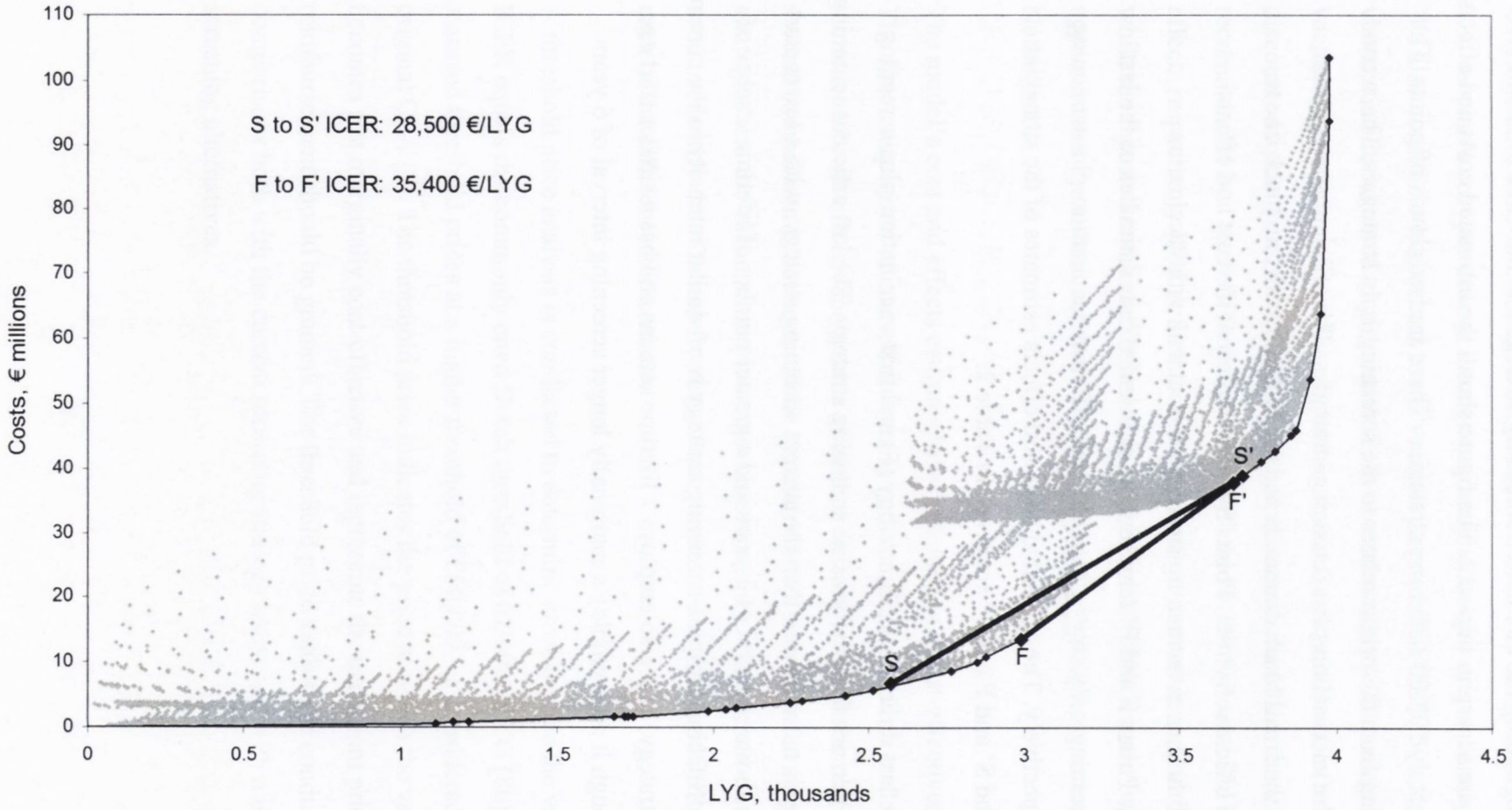
A threshold price analysis is conducted to determine at what price the vaccine's ICER equals the commonly cited Dutch threshold of €20,000/LYG [80]. We also assessed threshold prices at a higher threshold of €50,000, as considered in the original CEA [7]. The threshold price indicates the price at which the vaccine becomes just marginally cost-effective and represents the maximum price at which reimbursement should be granted. The threshold price analysis is conducted for comparisons both with the current screening strategy alone and with multiple screening alternatives.

RESULTS

The results of adding the alternative screening strategies to the CEA model of HPV vaccination are shown in Figure 2. The figure shows the estimated costs and effects of approximately 10,000 screening strategies. There are two clouds of points. The first spreading from the origin relates to the screening-only strategies. The second represents the vaccination-plus-screening strategies. The efficient strategies are shown with the small black diamonds and are joined by the thin black line to describe the efficient frontier. Points S and S' represent the cost and effects estimates of the current screening strategy without and with vaccination respectively. Points F and F' represent the cost and effects estimates of the last efficient screening-only strategy and the first efficient vaccination-plus-screening strategy respectively. The estimated costs and effects estimates of the strategies marked S and S' and F and F' are given in Table 2.

Notably, the last screening-only strategy (F) and first vaccination-plus-screening strategy (F') do not feature the same screening strategy. The last efficient screening-only strategy is more intense than the current screening strategy, with a lower start age of 28, a shorter interval of 3 years and a greater number of lifetime screens of 15. The first vaccination-plus-screening strategy is of similar intensity to the current screening strategy, as it also employs 7 lifetime screens and has a similar start age of 31, although it does employ a marginally longer screening interval of 6 years.

Figure 2: Cost-effectiveness plane depicting cost and effects of screening-only and vaccination-plus-screening alternatives illustrating the efficient frontier and the comparison between screening-only and vaccination-plus-screening when screening remain unchanged



The cord S S' represents the cost-effectiveness ratio of adding vaccination to the current screening strategy
The cord F F' represents the cost-effectiveness ratio between the last efficient screening-only strategy and the first efficient vaccination-plus screening strategy

Table 2: Screen schedule, cost, effects and ICERs of current screening without and with vaccination (S and S') and the last efficient screening-only strategy and the first efficient vaccination-plus-screening strategy (F and F') for a cohort of 100,000 women

Strategy	Screening interval	Start age	Stop age	Effects, LYG	Costs, €	ICER, €/LYG
S	5	30	60	2,570	6,712,000	
S'	5	30	60	3,700	38,925,000	28,500
F	3	28	70	2,990	13,654,000	
F'	6	31	67	3,680	37,825,000	35,400

The thick black line between S and S' represents the ICER of adding vaccination to the current screening intensity, which is €28,500/LYG. In comparison, the ICER between the last screening-only and first vaccination plus screening strategy, represented by the line F F', is €35,400/LYG. That is, assuming the current screening strategy will be maintained results in an ICER approximately €7,000/LYG less or 20% lower than the ICER when alternative screening intensities are considered.

Greater divergence between the ICERs is estimated at a higher vaccine price of €200 a dose, with an ICER of €46,800/LYG under the assumption of current screening intensity, compared to €62,000/LYG when alternative screening strategies are considered. Conversely, the divergence is much reduced and rank order of the ICERs reversed at a lower vaccine price of €50 a dose, with an ICER of €13,300/LYG under the assumption of current screening intensity, compared to €12,300/LYG when alternative screening strategies are considered. The threshold price analysis yields a vaccine price of €80 a dose assuming current screening only. Considering alternative screening strategies results in a marginally lower threshold price of €74.25. Greater divergence between the threshold prices is observed at a higher threshold of €50,000/LYG, with a threshold price of €214 per dose assuming current screening only and €163 per dose when alternative screening strategies are considered.

DISCUSSION

Interpreting the Results and Considering their Policy Relevance

The results of extending de Kok et al.'s original model to include alternative screening intensities shows that assuming screening remains fixed can exclude other more efficient strategies from the analysis. One consequence of an expanded choice set in this example is that the estimated ICER is less favourable than when screening is assumed unchanged. The ICER of the first vaccinated-plus-screening strategy of €35,400/LYG is nearly 25% greater than the ICER of adding vaccination the current screening strategy of €28,500/LYG. While this difference between the two cost-effectiveness ratios is not very large, it is not trivially small either.

As the current screening strategy without and with vaccination (points S and S') both lie very close to the efficient frontier, both can be effectively considered efficient. However, the presence of screening-only strategies to the south-east of the line S S' shows that at certain cost-effectiveness thresholds there are more efficient strategies than S or S'. This means that, starting from the current screening strategy without vaccination at point S, as the threshold increases it would be more efficient to increase the intensity of screening among the screening-only options than to switch to vaccination-plus-screening. Only once the threshold exceeds €35,400 would it be efficient to switch from a screening-only to a vaccination-plus-screening strategy.

The finding that extending the analysis to consider other screening alternatives does not lead to a large difference in the threshold price is a reassuring conclusion from a policy perspective. It indicates that the original CEA's assumption that screening remains unchanged would not lead to a large over-estimation of the appropriate vaccine price. Moreover, the fact that the difference between the ICERs falls as the vaccine's price is reduced is further reassurance considering the actual price paid for the vaccine in the Dutch vaccination programme is likely to be considerably less than the price of €118 per dose used in this analysis.

While the model suggests the policy significance of considering alternative screening strategies is modest in this case, it might not necessarily be so in all

circumstances. As the results show the difference between ICERs becomes much more pronounced at a higher vaccine price. Consequently, if the manufacturer sought a higher price for the vaccine, then the assumption that current screening is maintained would have led to a greater difference in the ICERs. Similarly, if the cost-effectiveness threshold was €50,000 not €20,000/LYG there would be a larger divergence between the threshold prices. Furthermore, if the vaccine is less effective than modelled, then the disparity between the ICERs will be greater. Indeed, it is possible that the vaccine may prove less effective than modelled here, as despite the fact that de Kok et al.'s original bases case analysis explicitly and deliberately employed a favourable set of assumptions for the vaccine's effectiveness, they note there remains considerable uncertainty regarding the vaccine's long term effectiveness.

The results show that the existence of potentially highly cost-effective secondary prevention in the form of screening limits the amount that can be efficiently charged for primary prevention through vaccination. Healthcare payers need to carefully consider the range of intervention options available when deciding if vaccination is cost-effective and at what price. By assuming screening remains fixed they ignore potential gains through further improvements to screening, which can influence the vaccine's cost-effectiveness and threshold price.

International Variation and Reasons for Restricted Comparisons

The literature review indicates that to date the majority of CEAs of HPV vaccination have assumed cervical screening remains fixed. It appears that proportionately more of the European analyses have only considered current screening to remain fixed than those analyses from the US or middle or low income countries. A possible explanation for this is that high income countries typically have established cervical screening programmes or widespread ad-hoc screening subject to screening guidelines. Consequently, decision makers are possibly inclined to request estimates of adding vaccination to the status quo, without considering alternative screening options. Conversely, low and middle income countries are less likely to have established screening provision, so decision makers considering cervical cancer prevention may be more likely to consider the possibilities for alternative screening strategies as well as vaccination.

It is notable that a number of US studies have considered alternative screening strategies. This might be a consequence of the fact that US cervical screening guidelines have recommended very intense screening, with short screening intervals and early start ages relative to programmes such as in the Netherlands. US analysts aware of the high intensity of screening may have wished to simulate a range of screening intensities to illustrate the scope for increasing efficiency by reducing screening intensity.

A further possible explanation for the majority of studies assuming screening remains constant is that modelling alternative screening strategies is more complex than assuming screening remains unchanged. Unless analysts have access to a previously established screening model they might be more inclined to consider current screening only out of convenience.

Some of the CEAs that only considered the current screening intensity noted it will be necessary to reconsider screening for vaccinated women [5, 14, 21, 29, 35]. The anticipated reduction in disease prevalence could justify a reduction in screening intensity or a switch to alternative testing methods, such as primary HPV testing [5]. However, these analyses consider changes to screening strategies to be a question for the future, not relevant for consideration now. Furthermore, they only consider the potential relevance of changing screening in vaccination-plus-screening strategies. However, as this analysis has demonstrated, considering screening alternatives is a relevant consideration now, as the screening-only strategies have relevance in determining the vaccine's cost-effectiveness.

Comparisons of screening programmes and guidelines in Europe indicates that Dutch screening is less intensive than in many other countries [81]. For example, the Dutch programme comprises 7 lifetime screens at intervals of 5 years from age 30, while German guidelines recommend annual screening from age 20 to beyond age 70. Our results indicate that difference in the ICERs between assuming screening remains unchanged and when alternative strategies are considered is likely to be greater when the initial screening intensity is lower. Therefore, the differences between the ICERs may not be as large in other countries, given the

higher baseline screening intensity. However, comparisons of screening coverage indicates that while the Netherlands achieves relatively high screening coverage, it can be much lower in other European countries [81]. Screening coverage may be even more relevant than screening intensity in countries with low levels of coverage for two reasons. Firstly, because increasing coverage is likely to be more cost-effective than adding vaccination when screening coverage is low (this is implied by Figure 2, as there are a large number of screening only strategies that lie below a line that could be drawn between the origin and any of the vaccination plus screening strategies. Secondly, the cost-effectiveness of adding vaccination will be more favourable when the screening coverage rate is low than when it is high.

Interpreting the ICER of Vaccination

Considering alternative screening strategies leads a methodological question of how to interpret the cost-effectiveness of the vaccine when implemented with screening. When screening is assumed constant the cost-effectiveness of vaccination is simply interpreted as the cost-effectiveness ratio of adding vaccination to the given screening programme. However, when estimating a complete cost-effectiveness frontier over multiple screening alternatives, there will only be one ICER on the frontier that compares a screening-only strategy and a vaccination-plus screening strategy (between F and F' in Figure 2). Since the last screening-only and first vaccination-plus-screening strategy do not necessarily feature the same screening strategy, there is no longer a simple ICER that can be reported as the cost-effectiveness ratio of adding vaccination to screening. Furthermore, the ICER between the last screening-only and first vaccination-plus-screening alternative will be the only ICER in the frontier responsive to vaccine price; although what strategies form these first and last strategies will vary with the vaccine price. Consequently, it may no longer be meaningful to report the cost-effectiveness of vaccination alone, but more useful to describe the cost-effectiveness of cervical cancer prevention strategies featuring vaccination.

The example of vaccination and screening provides an illustration of the interaction of two interventions for the prevention of the same disease. The example is particularly interesting because of the differing nature of the two interventions. Screening intensity is a key determinant of screening cost-effectiveness, which can

vary widely. HPV vaccination remains patent-protected so the intervention price is a variable subject to negotiation. Both the variables of screening intensity and price need to be considered when finding the most efficient strategy for a given cost-effectiveness threshold.

Limitations and Caveats

The model used in this analysis is a static model that does not capture the herd immunity effects of vaccination, which can lead to a conservative estimate of the vaccine's effectiveness. Similarly, the model also provides a somewhat conservative assessment of the vaccination's effects as it does not consider non-cervical HPV related disease such as penile, anal and head and neck cancer that have been assessed elsewhere [82]. However, the model's conservative assumptions may be offset in part by its favourable assumptions of lifelong protection by the vaccine, absence of side-effects and maintained screening adherence among vaccinated women.

The model presented here is deliberately specified to closely correspond to de Kok et al.'s original analysis. This is done to ensure the model used corresponds with policy alternatives available to decision makers at the time of the original analysis. We do not include alternative screening technologies such as liquid based cytology or HPV testing that have been considered since [83-85]. However, varying the screening intensity is sufficient to demonstrate the relevance of alternative screening strategies.

This paper does not feature an uncertainty analysis. While an assessment of uncertainty using a probabilistic sensitivity analysis (PSA) would be desirable, such an exploration of uncertainty in the present model was not possible, as the MISCAN model is not configured for PSA due to the computational burden of such an analysis in a microsimulation model.

One concern regarding the lack of a PSA is that the implied inefficiency of not considering alternative screening strategies could be small relative to any uncertainty around the efficient frontier revealed by a PSA. While a valid concern, there are a number of arguments as to why the analysis presented above is

appropriate for the primary purpose of this paper, which is to demonstrate the methodological consequences of omitting relevant comparators in a CEA of HPV vaccination.

While a PSA may result in a large scatter of estimates around the efficient frontier, it is possible that the inefficiency resulting from omitting alternative screening interventions would persist in many of these iterations. Furthermore, while the inefficiency caused by omitting alternatives may be eliminated in many PSA iterations, it is the averages of all PSA iterations which are relevant since expected costs and effects are the appropriate basis for decision making [86]. Although the results presented here are for a base-case parameter set rather than the mean estimates over multiple PSA iterations, it certainly seems plausible that the relevance of alternative screening strategies demonstrated here would probably also apply to the mean estimates from a PSA.

A further point justifying the analysis presented here is that while the implied inefficiency of omitting alternative strategies may only be modest, this could not be anticipated prior to the analysis: it was only by conducting the analysis that it could be shown that there was an implied efficiency loss and its size determined. Similarly, the modest efficiency loss identified in this study is contingent on a number of factors, including the current Dutch screening strategy and the vaccine price.

CONCLUSION

This analysis shows that modelling HPV vaccination in the Netherlands with the current screening intensity alone, rather than including alternative screening strategies led to a more favourable cost-effectiveness estimate of vaccination. Similarly, restricting the analysis to the current screening strategy results in a higher threshold price. The general conclusion is that alternative screening strategies are relevant for the estimation of vaccine cost-effectiveness both screening-only and vaccination-plus-screening strategies.

REFERENCES

1. Beutels P, Jit M. A brief history of economic evaluation for human papillomavirus vaccination policy. Sexual Health. 2010;7(3):352-8.

2. Barnabas R, Kulasingam SL. Economic evaluations of human papillomavirus vaccines. Expert Review of Pharmacoeconomics and Outcomes Research. 2007;7(3):251-67.

3. Torrance GW, Siegel JE, Luce BR. Framing and Designing the Cost-Effectiveness Analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 54-81.

4. Goldie SJ, Kohli M, Grima D, Weinstein MC, Wright TC, Bosch FX, et al. Projected Clinical Benefits and Cost-effectiveness of a Human Papillomavirus 16/18 Vaccine. Journal of the National Cancer Institute. 2004;96(8):604-15.

5. Rogoza RM, Westra TA, Ferko N, Tamminga JJ, Drummond MF, Daemen T, et al. Cost-effectiveness of prophylactic vaccination against human papillomavirus 16/18 for the prevention of cervical cancer: Adaptation of an existing cohort model to the situation in the Netherlands. Vaccine. 2009;27(35):4776-83.

6. Nielsen L. Classification of Countries Based on Their Level of Development: How it is Done and How it Could be Done. Washington: International Monetary Fund, 2011.

7. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of Human Papillomavirus vaccination in the Netherlands. Journal of the National Cancer Institute. 2009;101(15):1083-92.

8. Anonychuk AM, Bauch CT, Merid MF, Van Kriekinge G, Demarteau N. A cost-utility analysis of cervical cancer vaccination in preadolescent Canadian females. BMC Public Health. 2009;9(401):1-13.

9. Brisson M, Van de Velde N, De Wals P, Boily M-C. The potential cost-effectiveness of prophylactic human papillomavirus vaccines in Canada. Vaccine. 2007;25(29):5399-408.

10. Chesson HW, Ekwueme DU, Saraiya M, Markowitz LE. Cost-effectiveness of human papillomavirus vaccination in the United States. Emerging Infectious Diseases. 2008;14(2):244-51.

11. Elbasha E, Dasbach E, Insinga R. Model for assessing human papillomavirus vaccination strategies. Emerging Infectious Diseases. 2007;13(1):28-41.

12. Elbasha E, Dasbach E, Insinga R. A Multi-Type HPV Transmission Model. Bulletin of Mathematical Biology. 2008;70(8):2126-76.

13. Elbasha EH, Dasbach EJ, Insinga RP, Haupt RM, Barr E. Age-Based Programs for Vaccination against HPV. Value in Health. 2009;12(5):697-707.

14. Kim JJ, Goldie SJ. Health and Economic Implications of HPV Vaccination in the United States. New England Journal of Medicine. 2008;359(8):821-32.

15. Annemans L, my V, Oyee J, Largeron N. Cost-Effectiveness Evaluation of a Quadrivalent Human Papillomavirus Vaccine in Belgium. Pharmacoeconomics. 2009;27(3):231-45.

16. Bergeron C, Largeron N, McAllister R, Mathevet P, Remy V. Cost-effectiveness analysis of the introduction of a quadrivalent human papillomavirus vaccine in France. International Journal of Technology Assessment in Health Care. 2008;24(1):10-9.

17. Bogaards JA, Coupé VMH, Meijer CJLM, Berkhof J. The clinical benefit and cost-effectiveness of human papillomavirus vaccination for adult women in the Netherlands. Vaccine. 2011;29(48):8929-36.

18. Boot HJ, Wallenburg I, de Melker HE, Mangen M-JM, Gerritsen AAM, van der Maas NA, et al. Assessing the introduction of universal human papillomavirus vaccination for preadolescent girls in The Netherlands. Vaccine. 2007;25(33):6245-56.

19. Coupé VMH, van Ginkel J, de Melker HE, Snijders PJF, Meijer CJLM, Berkhof J. HPV16/18 vaccination to prevent cervical cancer in The Netherlands: Model-based cost-effectiveness. International Journal of Cancer. 2009;124(4):970-8.

20. Dasbach EJ, Insinga RP, Elbasha EH. The epidemiological and economic impact of a quadrivalent human papillomavirus vaccine (6/11/16/18) in the UK. BJOG: An International Journal of Obstetrics & Gynaecology. 2008;115(8):947-56.

21. Dasbach EJ, Largeron N, Elbasha EH. Assessment of the cost effectiveness of a quadrivalent HPV vaccine in Norway using a dynamic transmission model. Expert Review of Pharmacoeconomics and Outcomes Research. 2008;8(5):491-500.

22. Dasbach EJ, Nagy L, Brandtmüller A, Elbasha EH. The cost effectiveness of a quadrivalent human papillomavirus vaccine (6/11/16/18) in Hungary. Journal of Medical Economics. 2010;13(1):110-8.

23. Dee A, Howell F. A cost-utility analysis of adding a bivalent or quadrivalent HPV vaccine to the Irish cervical screening programme. The European Journal of Public Health. 2010;20(2):213-9.

24. Demarteau N, Detournay B, Tehard B, El Hasnaoui A, Standaert B. A generally applicable cost-effectiveness model for the evaluation of vaccines against cervical cancer. International Journal of Public Health. 2011;56(2):153-62.

25. Hillemanns P, Petry K, Largeron N, McAllister R, Tolley K, Büsch K. Cost-effectiveness of a tetravalent human papillomavirus vaccine in Germany. Journal of Public Health. 2009;17(2):77-86.

26. Jit M, Chapman R, Hughes O, Choi YH. Comparing bivalent and quadrivalent human papillomavirus vaccines: economic evaluation based on transmission model. British Medical Journal. 2011;343:1-15.

27. Lee VJ, Tay SK, Teoh YL, Tok MY. Cost-effectiveness of different human papillomavirus vaccines in Singapore. BMC Public Health. 2011;11(203):2-11.

28. Sanders GD, Taira AV. Cost-effectiveness of a potential vaccine for human papillomavirus. Emerging Infectious Diseases. 2003;9(1):37-48.

29. Taira AV, Neukermans CP, Sanders GD. Evaluating human papillomavirus vaccination programs. Emerging Infectious Diseases. 2004;10(11):1915-23.

30. Chen M-K, Hung H-F, Duffy S, Yen AM-F, Chen H-H. Cost-effectiveness analysis for Pap smear screening and human papillomavirus DNA testing and vaccination. Journal of Evaluation in Clinical Practice. 2011;17(6):1050-8.

31. Colantonio L, Gómez JA, Demarteau N, Standaert B, Pichón-Rivièrea A, Augustovski F. Cost-effectiveness analysis of a cervical cancer vaccine in five Latin American countries. Vaccine. 2009;27(40):5519-29.

32. Dasbach E, Insinga R, Yang Y, Pwu RF, Lac C, EH. E. The cost-effectiveness of a quadrivalent human papillomavirus vaccine in Taiwan. . Asia-Pacific Journal of Cancer Prevention. 2008;9(3):459-66.

33. Ezat WPS, Aljunid S. Cost-Effectiveness of HPV Vaccination in the Prevention of Cervical Cancer in Malaysia. Asia-Pacific Journal of Cancer Prevention. 2010;11(1):79-90.

34. Insinga RP, Dasbach EJ, Elbasha EH, Puig A, Reynales-Shigematsu LM. Cost-effectiveness of quadrivalent human papillomavirus (HPV) vaccination in Mexico: A transmission dynamic model-based evaluation. Vaccine. 2007;26(1):128-39.

35. Jit M, Choi YH, Edmunds WJ. Economic evaluation of human papillomavirus vaccination in the United Kingdom. British Medical Journal. 2008;337:1-12.

36. Kulasingam S, Benard S, Barnabas R, Largeron N, Myers E. Adding a quadrivalent human papillomavirus vaccine to the UK cervical cancer screening programme: A cost-effectiveness analysis. Cost Effectiveness and Resource Allocation. 2008;6(4):1-11.

37. La Torre G, de Waure C, Chiaradia G, Mannocci A, Capri S, Ricciardi W. The Health Technology Assessment of bivalent HPV vaccine Cervarix® in Italy. Vaccine. 2010;28(19):3379-84.

38. Liu PH, Hu FC, Lee PI, Chow SN, Huang CW, Wang JD. Cost-effectiveness of human papillomavirus vaccination for prevention of cervical cancer in Taiwan. BMC Health Services Research. 2010;10(11):1-11.

39. Mennini FS, Giorgi Rossi P, Palazzo F, Largeron N. Health and economic impact associated with a quadrivalent HPV vaccine in Italy. Gynecologic Oncology. 2009;112(2):370-6.

40. Obradovic M, Mrhar A, Kos M. Cost-effectiveness analysis of HPV vaccination alongside cervical cancer screening programme in Slovenia. The European Journal of Public Health. 2010;20(4):415-21.

41. Oddsson K, Johannsson J, Asgeirsdottir TL, Gudnason T. Cost-effectiveness of human papilloma virus vaccination in Iceland. Acta Obstetricia et Gynecologica Scandinavica. 2009;88(12):1411-6.

42. Olsen J, Jepsen MR. Human papillomavirus transmission and cost-effectiveness of introducing quadrivalent HPV vaccination in Denmark. International Journal of Technology Assessment in Health Care. 2010;26(2):183-91.

43. Sinanovic E, Moodley J, Barone MA, Mall S, Cleary S, Harries J. The potential cost-effectiveness of adding a human papillomavirus vaccine to the cervical cancer screening programme in South Africa. Vaccine. 2009;27(44):6196-202.

44. Suárez E, Smith JS, Bosch FX, Nieminen P, Chen C-J, Torvinen S, et al. Cost-effectiveness of vaccination against cervical cancer: A multi-regional analysis assessing the impact of vaccine characteristics and alternative vaccination scenarios. Vaccine. 2008;26(S5):F29-F45.

45. Szucs TD, Largeron N, Dedes KJ, Rafia R, Bénard S. Cost-effectiveness analysis of adding a quadrivalent HPV vaccine to the cervical cancer screening programme in Switzerland. Current Medical Research and Opinion. 2008;24(5):1473-83.

46. Thiry N, De Laet C, Hulstaert F, Neyt M, Huybrechts M, Cleemput I. Cost-effectiveness of human papillomavirus vaccination in Belgium: Do not forget about cervical cancer screening. International Journal of Technology Assessment in Health Care. 2009;25(2):161-70.

47. Torvinen S, Nieminen P, Lehtinen M, Paavonen J, Demarteau N, Hahl J. Cost effectiveness of prophylactic HPV 16/18 vaccination in Finland: results from a modelling exercise. Journal of Medical Economics. 2010;13(2):284-94.

48. Usher C, Tilson L, Olsen J, Jepsen M, Walsh C, Barry M. Cost-effectiveness of human papillomavirus vaccine in reducing the risk of cervical cancer in Ireland due to HPV types 16 and 18 using a transmission dynamic model. Vaccine. 2008;26(44):5654-61.

49. Vanagas G, Padaiga Ž, Kurtinaitis J, Logminienė Ž. Cost-effectiveness of 12- and 15-year-old girls' human papillomavirus 16/18 population-based vaccination programmes in Lithuania. Scandinavian Journal of Public Health. 2010;38(6):639-47.

50. Zechmeister I, de Blasio BF, Garnett G, Neilson AR, Siebert U. Cost-effectiveness analysis of human papillomavirus-vaccination programs to prevent cervical cancer in Austria. Vaccine. 2009;27(37):5133-41.

51. Kulasingam S, Connelly L, Conway E, Hocking JS, Myers E, Regan DG, et al. A cost-effectiveness analysis of adding a human papillomavirus vaccine to the Australian National Cervical Cancer Screening Program. Sexual Health. 2007;4(3):165-75.

52. Rogoza RM, Ferko N, Bentley J, Meijer CJLM, Berkhof J, Wang K-L, et al. Optimization of primary and secondary cervical cancer prevention strategies in an era of cervical cancer vaccination: A multi-regional health economic analysis. Vaccine. 2008;26(S5):F46-F58.

53. Sopina E, Ashton T. Cost-effectiveness of a cervical screening program with human papillomavirus vaccine. International Journal of Technology Assessment in Health Care. 2011;27(4):290-7.

54. Kim JJ, Ortendahl J, Goldie SJ. Cost-effectiveness of human papillomavirus vaccination and cervical cancer screening in women older than 30 years in the United States. Annals of internal medicine. 2009;151(8):538-45.

55. Kulasingam SL, Myers ER. Potential health and economic impact of adding a human papillomavirus vaccine to screening programs. Journal of the American Medical Association. 2003;290(6):781-9.

56. Tully SP, Anonychuk AM, Maria Sanchez D, Galvani AP, Bauch CT. Time for change? An economic evaluation of integrated cervical screening and HPV immunization programs in Canada. Vaccine. 2012;30(2):425-35.

57. Accetta G, Biggeri A, Carreras G, Lippi G, Carozzi FM, Confortini M, et al. Is human papillomavirus screening preferable to current policies in vaccinated and unvaccinated women? A cost-effectiveness analysis. Journal of Medical Screening. 2010;17(4):181-9.

58. Diaz M, de Sanjose S, Ortendahl J, O'Shea M, Goldie SJ, Bosch FX, et al. Cost-effectiveness of human papillomavirus vaccination and screening in Spain. European Journal of Cancer. 2010;46(16):2973-85.

59. Campos NG, Kim JJ, Castle PE, Ortendahl JD, O'Shea M, Diaz M, et al. Health and economic impact of HPV 16/18 vaccination and cervical cancer screening in Eastern Africa. International Journal of Cancer. 2012;130(11):2672-84.

60. Canfell K, Shi J-F, Lew J-B, Walker R, Zhao F-H, Simonella L, et al. Prevention of cervical cancer in rural China: Evaluation of HPV vaccination and primary HPV screening strategies. Vaccine. 2011;29(13):2487-94.

61. Diaz M, Kim JJ, Albero G, de Sanjose S, Clifford G, Bosch FX, et al. Health and economic impact of HPV 16 and 18 vaccination and cervical cancer screening in India. British Journal of Cancer. 2008;99(2):230-8.

62. Goldie SJ, Diaz M, Kim S-Y, Levin CE, Van Minh H, Kim JJ. Mathematical Models of Cervical Cancer Prevention in the Asia Pacific Region. Vaccine. 2008;26(S12):M17-M29.

63. Goldie SJ, Diaz M, Constenla D, Alvis N, Andrus JK, Kim S-Y. Mathematical Models of Cervical Cancer Prevention in Latin America and the Caribbean. Vaccine. 2008;26(S11):L59-L72.

64. Goldie SJ, Kim JJ, Kobus K, Goldhaber-Fiebert JD, Salomon J, O'Shea MKH, et al. Cost-effectiveness of HPV 16, 18 vaccination in Brazil. Vaccine. 2007;25(33):6257-70.

65. Kim JJ, Kobus KE, Diaz M, O'Shea M, Van Minh H, Goldie SJ. Exploring the cost-effectiveness of HPV vaccination in Vietnam: Insights for evidence-based cervical cancer prevention policy. Vaccine. 2008;26(32):4015-24.

66. Praditsitthikorn N, Teerawattananon Y, Tantivess S, Limwattananon S, Riewpaiboon A, Chichareon S, et al. Economic Evaluation of Policy Options for Prevention and Control of Cervical Cancer in Thailand. Pharmacoeconomics. 2011;29(9):781-806.

67. Reynales-Shigematsu LM, Rodrigues ER, Lazcano-Ponce E. Cost-Effectiveness Analysis of a Quadrivalent Human Papilloma Virus Vaccine in Mexico. Archives of Medical Research. 2009;40(6):503-13.

68. Sharma M, Ortendahl J, van der Ham E, Sy S, Kim JJ. Cost-effectiveness of human papillomavirus vaccination and cervical cancer screening in Thailand. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(2):166-76.

69. Ginsberg GM, Edejer TT-T, Lauer JA, Sepulveda C. Screening, prevention and treatment of cervical cancer: A global and regional generalized cost-effectiveness analysis. Vaccine. 2009;27(43):6060-79.

70. Ginsberg GM, Fisher M, Ben-Shahar I, Bornstein J. Cost-utility analysis of vaccination against HPV in Israel. Vaccine. 2007;25(37-38):6677-91.

71. Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ. Cost-effectiveness of cervical cancer screening with Human Papillomavirus DNA testing and HPV-16,18 Vaccination. Journal of the National Cancer Institute. 2008;100(5):308-20.

72. Yamamoto N, Mori R, Jacklin P, Osuga Y, Kawana K, Shibuya K, et al. Introducing HPV vaccine and scaling up screening procedures to prevent deaths from cervical cancer in Japan: a cost-effectiveness analysis. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(2):177-86.

73. Goldie SJ, O'Shea M, Campos NG, Diaz M, Sweet S, Kim S-Y. Health and economic outcomes of HPV 16,18 vaccination in 72 GAVI-eligible countries. Vaccine. 2008;26(32):4080-93.

74. Habbema JDF, van Oortmarssen GJ, Lubbe JTN, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine. 1985;20(1):79-93.

75. van Ballegooijen M. Effects and costs of cervical cancer screening. Rotterdam: Erasmus University; 1998.

76. Nederlandse Vereniging voor Pathologie (Dutch Society for Pathology). PALGA: The Nationwide Network and Registry of Histo- and Cytopathology in the Netherlands.

77. Jacobs MV, Walboomers JMM, Snijders PJF, Voorhorst FJ, Verheijen RHM, Fransen-Daalmeijer N, et al. Distribution of 37 mucosotropic HPV types in women with cytologically normal cervical smears: The age-related patterns for high-risk and low-risk types. International Journal of Cancer. 2000;87(2):221-7.

78. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JDF. Cost-effectiveness of cervical cancer screening: Comparison of screening policies. Journal of the National Cancer Institute. 2002;94(3):193-204.

79. CVZ. Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie [in Dutch]. Diemen: College voor Zorgverzekeringen, 2006.

80. Boersma C, Broere A, Postma MJ. Quantification of the Potential Impact of Cost-effectiveness Thresholds on Dutch Drug Expenditures Using Retrospective Analysis. Value in Health. 2010;13(6):853-6.

81. Anttila A, von Karsa L, Aasmaa A, Fender M, Patnick J, Rebolj M, et al. Cervical cancer screening policies and coverage in Europe. European Journal of Cancer. 2009;45(15):2649-58.

82. de Kok IMCM, Habbema JDF, van Rosmalen J, van Ballegooijen M. Would the effect of HPV vaccination on non-cervical HPV-positive cancers make the difference for its cost-effectiveness? European Journal of Cancer. 2010;47(3):428-35.

83. van Rosmalen J, de Kok IMCM, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(6):699-709.

84. de Bekker-Grob E, de Kok I, Bulten J, van Rosmalen J, Vedder J, Arbyn M, et al. Liquid-based cervical cytology using ThinPrep technology: weighing the pros and cons in a cost-effectiveness analysis. Cancer Causes and Control. 2012;23(8):1323-31.

85. de Kok IMCM, van Rosmalen J, Dillner J, Arbyn M, Sasieni P, Iftner T, et al. Primary screening for human papillomavirus compared with cytology screening for cervical cancer in European settings: cost effectiveness analysis based on a Dutch microsimulation model. British Medical Journal. 2012;344(7847):1-17.

86. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. Pharmacoeconomics. 2008;26(9):781-98.

SUMMARY & CONCLUSION

This chapter briefly reviews the aims of the studies within the thesis, explains how these were met and summarises what each study contributes to our understanding of CEA. This chapter describes the benefit to decision makers of each of these enhancements to our understanding of CEA explains how each study relates to the thesis's central theme of how CEAs can be reframed to better correspond to the policy questions they are to inform. In addition, it also notes some of the caveats and limitations of each of the studies and describes potential topics for further research arising from each. Finally, this chapter closes with some concluding observations on some of the general issues encountered during the course of the research for the thesis and some thoughts on how the fit between CEAs and policy questions might be improved in the future.

Paper 1: The Problem of Comparability

Aims & Contribution

The aim of the first paper was to provide a simple demonstration of the problem of comparability between CEAs employing different numbers of future cohorts when differential discounting of costs and health effects is applied. The demonstration of the problem of comparability contributes to our understanding of CEA by showing how differential discounting can distort the choice of optimal intervention. This contribution to the current literature was necessary as there was no other study that addressed this particular topic or showed the consequences of differential discounting for comparisons between CEAs.

Benefit to Decision Makers

Awareness of the problem of comparability will aid decision makers in reaching the appropriate conclusions regarding the cost-effectiveness of interventions. An understanding of the problem of comparability informs decision makers that they cannot directly compare ICERs from interventions with different numbers of future cohorts, without considering the effect of the increased willingness to pay over time implied by differential discounting. An awareness of the problem of comparability

will also help decision makers avoid accepting CEAs that have strategically included a large number of future cohorts to achieve low ICERs.

Awareness of the problem of comparability will also aid those drafting CEA guideline documents, as they will now have an understanding of the potential difficulties when recommending differential discounting. Indeed, awareness of the problem of comparability places an onus upon on those recommending differential discounting to provide sufficient guidance regarding is appropriate application. Furthermore, an understanding of the problem comparability is necessary when considering how best to solve it.

Implications for Model Framing

Prior to the adoption of differential discounting CEA models were, in a manner, time-neutral. It did not matter if the intervention was modelled in a cohort now, or in the future, or indeed over multiple cohorts both now and in the future. In this regard framing the question in terms of time and the number of recipient cohorts was not important. However, with the application of differential discounting, the temporal position of the recipients now matters.

Since an intervention's recipients' position in time matters under differential discounting, we must now consider if we wish to estimate cost-effectiveness just for the current cohort alone or do we wish to consider future cohorts too. Consequently, applying differential discounting requires us to think about the temporal framing of our model.

Caveats and Limitations

Paper 1 employs a static model, meaning the herd immunity effects HPV vaccination are not modelled. This aspect of the paper drew comment in letter by Westra et al. [1]. As noted in the reply to Westra et al. [2], while it is true that a dynamic model would be more appropriate for the assessment HPV vaccination, the purpose of the analysis was to demonstrate the effect of differential discounting in multiple cohorts. A static model was convenient in this case, as it ensured the cost-effectiveness of vaccination improved only with the inclusion of future cohorts, rather than with the accumulation of indirect effects of vaccination.

A limitation of Paper 1 is that it is only relevant to the few countries that currently apply differential discounting. Furthermore, the extent to which ICERs fall as new cohorts are added to a CEA depends on the size of the discounting differential. The smaller the differential, the less pronounced the problem of comparability will be. Since arguably discounting differentials should be smaller than they are now, the problem of comparability might be much smaller if more modest discounting differentials were adopted. However, this limitation only applies in principle, as in practice the loss of comparability can matter due to the relatively large differentials currently implemented.

Further Research

A topic of interest for further research arising from the first paper would be a consideration of the discounting differential used in the Netherlands. The current Dutch differential of 2.5% was not set with reference to expected threshold growth. Indeed, the Netherlands has no official cost-effectiveness threshold. An analysis of the cost-effectiveness of interventions accepted and rejected in the Netherlands could however reveal an implicit threshold and provide evidence of growth over time or otherwise. If evidence of threshold growth weaker than 2.5% per annum is found this would then require a reconsideration of Dutch discount rates, which could have profound implications for the cost-effectiveness of preventative interventions such as HPV vaccination.

Paper 2: Solving the Problem of Comparability

Aims & Contribution

The publication of Paper 1 demonstrating the problem of comparability left an obvious gap in the literature for a method proposing a solution. The aim of the second paper was to resolve the problem of comparability, allowing the direct comparison of cost-effectiveness estimates between multi-cohort models under differential discounting. Paper 2 provides an intuitive solution to the problem of comparability. The proposed solution allows CEAs from multi-cohort models to be compared to other studies and to the current cost-effectiveness threshold under differential discounting.

Benefit to Decision Makers

Differential discounting is an attempt to achieve greater correspondence between models and policy choices, given the anticipation of threshold growth. While the intra-cohort effect of differential discounting described in Paper 2 is intended, the inter-cohort effect is not and leads to the problem of comparability. The proposed adjustment of the hypothetical comparator allows the original intent of differential discounting to be realised by preserving the intra-cohort effect, but eliminates the inter-cohort effect, thus ensuring greater correspondence between the model and real policy choices.

The proposed solution will be of benefit to decision makers unsure of how to compare cost-effectiveness estimates due to the problem of comparability. In particular, the application of the adjusted ICER allows decision makers to compare the cost-effectiveness of interventions in the manner that they are currently accustomed, without having to give any special consideration for the number of cohorts modelled.

Relevance to Model Framing

In practice, what the hypothetical comparator allows us to do is apply differential discounting without worrying about the temporal framing of the policy question. The hypothetical comparator means the inclusion of one or many future cohorts no longer matters under differential discounting. Consequently, the time-neutrality that

applied under equal discounting is restored. As such, CEA practitioners need not worry about the intertemporal framing of their model and its correspondence to the policy question at hand.

Caveats and Limitations

The first caveat to note is that although the hypothetical comparator allows us to overcome the problem of comparability, it adds another layer of complexity to CEA. Anything that makes modelling more opaque may have the effect of reducing its impact on decision makers.

A second point of note is that the hypothetical comparator is only required in the rather specific circumstances of analyses employing differential discounting that feature cohorts starting the intervention after the discount year. Since few countries currently employ differential discounting, the number of analyses presently requiring such an adjustment will be small.

A related issue is that if the discounting differential is determined by estimates of threshold growth as implied by Claxton et al.'s analysis [3], then the differential is likely to be small, as plausible rates of threshold growth are low. The lower the differential, the smaller the distortion caused by the inter-cohort effect and the more modest the hypothetical comparator's adjustment factor. Similarly, the fewer future cohorts modelled, the smaller the implied adjustment factor. In cases in which the adjustment factor is small, it may be considered acceptable not to apply it, as the unadjusted estimates will closely approximate the adjusted estimates.

A final point of note is that the analysis of Paper 2 does not consider the effect of uncertainty. Uncertainty is generally of limited relevance to this study, as the need for a correction for the number of future cohorts applies irrespective of any uncertainty in cost-effectiveness estimates and the inclusion of an uncertainty analysis would not alter the study's conclusions. Indeed, it is important to note that the proposed method of the hypothetical comparator could be useful in ameliorating the effects of uncertainty if the discounting differential was modelled as a random variable. In principle, it is appropriate to model the discounting differential as a random variable, as it is not known with certainty, since growth in the cost-

effectiveness threshold, which in principle should inform the size of the discounting differential, is also uncertain. A model with a randomly varying discounting differential will produce more uncertain results without the application of the hypothetical comparator advocated in Paper 2 than with it.

Further Research

The example of the cost-effectiveness of adding HPV vaccination to current screening improving with the inclusion of future cohorts under differential discounting is a relatively simple one. However, what is more complex is the consideration of how a complete cost-effectiveness frontier of an intervention with multiple cost and effects combinations such as screening changes over time with the inclusion of new cohorts under differential discounting. Research into this topic would consider whether the hypothetical comparator should be applied to such a case or whether some alternative means for correcting for the inter-cohort effect should be employed.

Paper 3: Implementation Time Horizons

Aims & Contribution

The aim of the third paper was to describe the implementation horizon and assess its importance to a CEA used to inform colorectal screening policy in the Netherlands. This aim was motivated by the lack of existing literature describing the implementation horizon or quantifying its effects in an applied example.

The primary contributions of this paper are the description of the implementation horizon and the demonstration that a 30-year horizon had no meaningful impact on optimal screening recommendations in the CEA considered. This study also adds to the literature by providing a survey of CEAs in the screening literature employing implementation horizons. Finally, a more general contribution of this paper is the novel presentation of the divergence of optimal intervention strategies between two models using the net benefit framework.

Benefit to Decision Makers

Paper 3 will be of assistance to decision makers by allowing them to recognise and describe an implementation horizon in any analysis they are presented with. The example of the comparison in the net benefit framework also provides a simple and easy to use template which other analysts might apply to assess the impact of a horizon in other applications.

The analysis of Paper 3 also is of benefit to those drafting CEA methods guidelines. It shows how there is more to time horizons than the horizon for the assessment of costs and health effects.

Framing the Model

The example presented in Paper 3 shows that reframing the analysis to consider the lifetime of screening for the simulated cohorts rather than over a finite period of 30 years does not meaningfully alter the implied optimal screening strategies. However, the analysis does usefully call attention to this aspect of model framing and provides us with the terminology to describe it and demonstrated how it might be appraised quantitatively.

Caveats and Limitations

The obvious limitation of Paper 3 is that it only considers one example. A possible extension to the analysis would be to consider screening for another disease, as factors such as screen sensitivity and length of preclinical duration may influence the effect of the implementation horizon. Another possible extension would be to investigate if the effects of the implementation horizon vary with prior screening history. For example, the Dutch colorectal CEA considers the introduction of screening in a population without previous organised screening. It is possible that the influence of the implementation horizon could moderate further if prevalent cohorts have some previous screening history.

An apparent limitation of Paper 3 is that it investigated a feature of modelling that although apparently unrepresentative of actually policy choices, did not result in any meaningful differences in the optimal policy choices when compared to a model more apparently representative of reality. As such, the study seems to present a

negative result in that the implementation horizon does not appear to matter. However, it is difficult to anticipate which model assumptions will matter and which will not. Only by undertaking the analysis without the implementation horizon was it revealed that reframing the model in this way does not meaningfully alter implied optimal policies. Furthermore, the analysis showed that a shorter horizon would meaningfully alter the optimal policy choices identified by the model. So, in principle the implementation horizon can be important and this is a useful result.

Further Research

The colorectal example considered in third paper featured a cost-effectiveness frontier without a pronounced curve: the highest ICERs in the curve were not very much greater than the lowest. By contrast, the cost-effectiveness frontier of screening only strategies in Paper 6 has a much more pronounced curve, featuring both modest and exceptionally high ICERs. A simple extension of Paper 3 would be to appraise the 27-year horizon used in van der Akker-van Marle et al.'s CEA of cervical screening to see if the loss of net benefit due to imposing an implementation horizon is similarly modest in the case of an intervention with a more pronounced curve in its cost-effectiveness frontier.

Paper 4: Aggregation in Multi-Cohort Models

Aims & Contribution

The existing literature that most explicitly addresses the multi-cohort model structure in CEA advocates the aggregation of cost-effectiveness estimates across cohorts. The aim of the fourth paper was to demonstrate that aggregating cost-effectiveness estimates over all cohorts in multi-cohort models does not lead to optimal policy choices.

The contribution of this paper is to provide a clear counterargument supported by an example to the existing publications on multi-cohort models that advocate reporting aggregate estimates over all cohorts. By providing an alternative perspective, this paper adds balance to the literature regarding the appropriate model structure for the simulation of interventions with multiple concurrent recipient cohorts.

Benefit to Decision Makers

By providing a counterargument to the view that cost-effectiveness estimates should be aggregated across cohorts Paper 4 provides decision makers with an alternative perspective from which to appraise the appropriateness of model structures in the CEAs presented to them. It also provides decision makers with a clear articulation of the potential problems of aggregation, which should encourage them to critically reflect on what model structure is most appropriate to the given policy question when commissioning a CEA.

CEA methods guidance statements such as those issued by NICE and the International Society for Pharmacoeconomic and Outcomes Research (ISPOR) and the Society of Medical Decision Making (SMDM) do acknowledge the principle of subgroup analysis, suggesting that separate cost-effectiveness estimates be reported for separate subgroups. However, they do not reflect on how the principle of subgroup analysis applies to model structure. While such methods guidance documents cannot plausibly provide an exhaustive description of every detail of appropriate model structure, they can reference to the relevant literature. Paper 4 provides a useful supporting description of the issue of aggregation that could be cited in any such methods guidance.

Framing the Model

The fundamental point behind Paper 4 is that the presumption should not be to frame a CEA around an intervention, but rather consideration should be given to whether the CEA should be framed around the intervention's recipients. This is relevant wherever the intervention can be implemented differently in different potential recipient groups and where cost-effectiveness may vary between such recipient groups. The analysis showed how such a reframing from the intervention to the intervention's recipients would change what cost-effectiveness estimates are reported and what are the implied reimbursement decisions.

Caveats and Limitations

Paper 4 notes how the ICERs and ACERs of the adoption of a HPV based test vary between cohorts in a multi-cohort model. These cost-effectiveness ratios are widely relied upon in CEA as the primary measures of cost-effectiveness, which is why Paper 4 employed them as the primary metric of cost-effectiveness. However, they are not useful for assessing how inefficient it would be to impose a common strategy over all cohorts. More insight could be offered by assessing the costs in terms of net benefit of aggregating cost-effectiveness estimates over cohorts, using the same framework as employed in Paper 3 An assessment of the loss of net benefit of imposing a common strategy over all cohorts could find that the loss of efficiency would only be modest.

The example used in Paper 4 could be perceived as somewhat contrived, as it includes cohorts that have received HPV vaccination. The anticipated reduction in disease prevalence in the vaccinated cohorts results in pronounced differences in cost-effectiveness between the aggregate and the per cohort estimates for those cohorts. In the unvaccinated cohorts the ICERs only vary with screening history and the differences are much smaller. In other screening examples without a radical change in anticipated disease prevalence such as that following HPV vaccination, the differences in ICERs between cohorts might also only be modest.

A limitation of the analysis presented in Paper 4 is that it only compares two screening tests without considering alternative screening intervals and start and stop ages. A more complete analysis would consider alternative schedules, such as that

by van Rosmalen et al. [4]. However, the current analysis is sufficient for the purposes of illustrating the potential problems of aggregating estimates over cohorts.

Further Research

One particular application of the idea of considering cohort-specific models to identify optimal screening would be to consider stopping ages for cervical cancer screening. The use of single cohort models in many cervical screening models to date means the optimal screening stop age is identified for simulated women with a prior screening history. However, the optimal screening stop age for women without a screening history may older. Given that individuals' health seeking behaviour may change over time, it would be useful to identify the optimal screening ages for older cohorts without prior screening histories.

Paper 5: Adjusting the Optimal Screening Interval for Disease Incidence

Aims & Contribution

Mathematical models of cost-effectiveness have the benefit that they may be solved to determine the relationship between parameters and outcome variables. However, mathematical models are now less common in the literature, as more analysts adopt simulation models instead. One consequence of the switch to simulation models is that attempts to illustrate the relationship between disease incidence and the optimal screening interval explored in earlier mathematical models have not been pursued in the recent literature. This left an obvious gap in the literature for studies considering this relationship with the benefit of modern modelling techniques.

Paper 5 revisits the mathematical approach and the incidence-optimal interval relationship considered previously by Kirch and Klein. The aim is to provide a simple model for identifying the change in the optimal screening interval for a given risk-subgroup relative to the optimal screening interval for the average risk group.

Paper 5 contributes to the literature by presenting a more detailed model than that of Kirch and Klein that also has the benefit of a validation against a modern

microsimulation model, which has previously been used to determine optimal screening intervals. As such, Paper 5 combines theoretical insights from simpler earlier models with validation against more sophisticated recent modelling approaches.

Benefit to Decision Makers

The development of simulation models for the assessment of screening strategies requires experienced modellers and a considerable amount of data. Therefore, it may be common for such modelling efforts to be made for the majority average risk group alone, leaving decision makers without a guide to optimal policy for specific risk subgroups. As a result, the recommendations for the average risk group may not adequately correspond to the optimal screening strategies for those of high disease risk.

The model presented in Paper 5 describes the relationship between cancer incidence and the optimal screening interval. A new model does not then need to be specified and run for each subgroup separately, but the risk-specific optimal interval can be found from the described relationship. As such, the model in Paper 5 provides decision makers with a relatively easy to implement tool for adjusting screening intervals for high risk groups. This may be of particular benefit to decision makers without ready access to a full microsimulation model of modelling expertise. This incidence-interval relationship can also be used to identify how the optimal grouping of risk subgroups, given the range of possible screening intensities.

Reframing the Analysis

The mathematical model demonstrated in Paper 5 shows how the question of optimal screening can be reframed from the conventional question of investigating what the optimal screening strategy is for a given group, to a broader, more abstract consideration of how optimal screening varies between subgroups. The ultimate goal of achieving the optimal strategy for each subgroup remains the same, but by changing the mode of analysis and using the simplified mathematical model we can arrive at answers faster and more efficiently than we might have otherwise.

Caveats and Limitations

The mathematical model employs a number of simplifying assumptions. Much of the relevant detail about costs and effects is sacrificed in order to achieve a simple, mathematically tractable model that still provides an adequate description of the change in the optimal interval with disease incidence. The validation exercise finds the model does not predict the same absolute size of optimal screening interval as the full microsimulation model, but does however provide a good fit for the relative change in the optimal screening interval with an increase in disease incidence from the average risk rate, although the fit is less satisfactory at incidence rates below average risk.

A limitation of the mathematical model is that it is highly abstracted and does not consider alternative screening start and stop ages, only varying intervals. It cannot be used to specify risk-specific screening age ranges.

A limitation of the model and the validation exercise is that while disease incidence is assumed to increase, the rate of disease progression is assumed constant in both the mathematical and microsimulation models. This assumption of constant disease progression may not be justified in all cases, as some high risk groups may be subject to faster disease progression.

It should be noted that there are two principal differences between the mathematical model and the MISCAN microsimulation model employed in Paper 5. The first is that the mathematical model employs mathematical methods rather than simulation and the second is that it is a highly simplified model that ignores much of the detail captured in the microsimulation model. Therefore, it should be recognised that much of the benefit of the mathematical model could be realised by employing a similarly simplified microsimulation model.

Employing a simplified microsimulation model rather than a simplified mathematical model may be more suitable, given that the main variables of screening costs and disease incidence may largely capture the relationship between the change in incidence and the optimal interval. Furthermore, the application of such a simplified mathematical model may be considerably more accessible than the

mathematics employed in the model in Paper 5. However, it should equally be noted that such simplified microsimulation model would not have the benefit of offering an analytic solution to the incidence-optimal interval relationship that is particular to the mathematical model.

Further Research

A useful extension of the model in Paper 5 would be to validate it in another disease case. A convenient application would be to again validate the model against the MISCAN microsimulation model, this time in the case of cervical screening. The model used in Paper 6 showing the optimal screening intervals in vaccinated and unvaccinated women could be used to provide an example of low and high risk women.

Paper 6: The Relevance of Alternative Screening Strategies to CEAs of HPV Vaccination

Aims & Contribution

To date there has been no study explicitly examining the implications of including alternative cervical screening strategies when assessing the cost-effectiveness of HPV vaccination. Paper 6 addresses that gap by surveying the literature to describe the approaches taken in published CEAs and by providing a quantitative assessment of the consequences of failing to consider alternative screening strategies using an applied example.

Paper 6 contributes to the literature by showing that alternatives screening strategies are relevant comparators when assessing the cost-effectiveness of HPV vaccination. While this is of use to modellers and decision makers in the specific field of cervical cancer prevention as they consider further enhancements to screening strategies and possible new polyvalent vaccines in the future, the example also provides a useful illustration of the principle of how the cost-effectiveness of one intervention may be contingent on the cost-effectiveness of another complementary intervention.

The principle that all relevant comparators be compared is well recognised in CEA and this is reflected in methods guidance documents. So the relevance of Paper 6 is

not with regard to the principle of including all relevant comparators, but rather that so many published CEAs failed to adhere to this well established principle when assessing the cost-effectiveness of HPV vaccination.

Benefit to Decision Makers

Paper 6 is not likely to be useful to decision makers considering HPV vaccination, as most countries have now adopted the intervention. However, the example provided by Paper 6 is a useful illustration that there may be no ICER of a given intervention in the conventional sense used in cost-effectiveness analysis, but rather a level of the cost-effectiveness threshold at which the intervention becomes cost-effective. This issue will primarily be of interest to those working the field of cost-effectiveness analysis theory rather than to applied decision makers. It is relevant to CEA analysts, who need to reflect on what cost-effectiveness evidence they present to decision makers.

Reframing the Question

Paper 6 offers a simple example of how easy it can be to unintentionally frame a policy question in a restrictive way that neglects to consider relevant policy alternatives. It is clear that considering the addition of HPV vaccination to current cervical screening services may seem an appropriate policy choice. However, it is only when the analysis is reframed to the broader perspective of what is the optimal combination of screening intensities and vaccination does it become clear that changing the screening intensity is a relevant consideration when appraising the cost-effectiveness of vaccination.

Caveats and Limitations

Like Paper 1, Paper 6 has the limitation that the model used is static, and so cannot simulate the indirect effects of vaccination. While the model could thus be interpreted as a conservative estimate of the vaccine's effectiveness, the model also includes a number of assumptions favourable to vaccination including lifelong protection, the absence of side-effects and no reduction in screening adherence [5], which may attenuate any underestimation of vaccine effectiveness.

While the model indicates a non-trivial difference in ICERs between the analyses without and with alternative screening strategies, inspection of the cost-effectiveness plane in Paper 6 indicates that the loss of net benefit (at a threshold at which vaccination would be cost-effective) of failing to consider alternative screening strategies is relatively modest compared to the total anticipated net benefit.

A further caveat is that the observed discrepancy between the ICERs when alternative screening strategies are and are not included depends on the current screening intensity. The Netherlands has a low screening intensity relative to most developed countries. The discrepancy between ICERs is likely to be less or possibly even reversed if the current screening intensity were higher.

Finally, the analysis in Paper 6 does not consider uncertainty. While it would be desirable to include a PSA, this is precluded by the technical limitation that the MISCAN microsimulation model is not currently configured for conducting a PSA due to the computational burden. The modest difference in net benefit resulting from the inclusion of alternative screening strategies may be small relative to uncertainty around the cost-effectiveness frontier. However, for the reasons described in the paper, the presence of uncertainty is not a reason to disregard the relevance of alternative strategies.

Further Research

The analysis presented in Paper 6 is a reflection on modelling methods used in the previous appraisals of HPV vaccination. The principle of the relevance of alternative screening strategies will again be pertinent with the likely development of improved HPV vaccines with protection against more HPV types. CEA comparisons of such polyvalent vaccines to current multivalent vaccines would also have to give consideration to alternative screening strategies. However, such further research would have to include cost-effectiveness estimates of new screening techniques such as HPV based screening, rather than just the conventional cytology considered in Paper 6.

Reframing and the Process of Methods Development

This thesis considers how CEA models can be reframed to better correspond to the policy questions they are to inform. The preceding section summarised the specific studies in the thesis. It reviewed examples in which aspects of modelling imply a divergence between models and the appropriate policy questions and also how these models can be reframed to eliminate this divergence. This section considers the issue of reframing in the broader context of CEA methods development, including the efforts currently being made to improve CEA methods and how they relate to the need for reframing addressed in this thesis.

The process of methods development has always been central to the field of CEA. This is evidenced by the numerous textbooks, methodological articles and methods guidelines published in the CEA literature. The methods guidance documents include consensus statements from expert panels drawn from the professional societies within the discipline, such as the recent guidelines jointly published by ISPOR and SMDM [6]. Statutory CEA authorities also publish methods guidelines, such as those produced by NICE in England and Wales, the Canadian Agency for Drugs and Technology in Health and the Health Information and Quality Authority in Ireland [7-10].

The need for appropriate framing has been recognised in the past and remains of interest to those currently writing on cost-effectiveness methods, as evidenced by Torrance et al. cited in the introduction to the thesis and more recent publications on the topic [11, 12]. However, much of the current efforts in methods development are directed at other topics, such as improving modelling techniques with the use of discrete event simulation instead of state transition modelling [13], better representations of uncertainty in CEA [14] and improvements to quality of life measures [15]. By contrast, the research undertaken as part of this thesis found the literature on aspects of model fit to policy questions to be limited. For example, there were only two existing papers that directly considered the arguments for multi-cohort modelling in the case of screening interventions and there was no pre-existing work on the question of implementation horizons. With relatively little research being undertaken on the adequacy of model fit, it is perhaps unsurprising that there remains scope for improvement in the framing of CEAs.

In addition to the lack of literature dedicated to considerations of model fit there is a second aspect to the process of methods development relevant to the question of model framing. In some cases the adoption of novel methods will lead to unanticipated and unintended consequences for model fit to policy choices. Consequently, there will be an ongoing need to consider the implications of such methodological innovations for model framing.

This thesis contains two relevant examples of methodological innovations that have entered practice which have implications for model framing. The case of differential discounting in multi-cohort models neatly exemplifies the adoption of a new method that resulted in unanticipated consequences: the problem of comparability described in Paper 1. Although not considered in one of the central papers of this thesis, a similar example of a novel method that had unanticipated consequences for the usefulness of CEA results for decision makers is described in two posters in the Appendix. Namely that cost-effectiveness acceptability curves (CEACs) are not suitable measures of uncertainty in the case of interventions in which there are many possible alternative strategies.

Although the question of appropriate framing does not appear to be a priority in current methods research and the adoption of novel methods itself continues to pose challenges for the appropriate framing of CEAs, the process of methods development should not be seen in tension with adequate model framing. Adequate model framing is just one part of good modelling practice. The discipline of CEA remains committed to such good practice, through the process of guideline revision mentioned above, which is complemented by the process of peer-review academic publication. Accordingly, the field of CEA should, in principle, be receptive to research regarding model framing. That relatively little has been published on framing probably simply reflects inattention to this aspect of modelling by researchers in the field to date.

Conclusion

The contribution of this thesis is an examination of how models can diverge from policy questions in ways that are relatively simple but not always obvious. It shows how such divergence can arise from the application of novel methods without their consequences being fully understood, or simply through the application of existing methods without sufficient consideration for the appropriate policy questions. The thesis uses a number of applied examples to show how CEA models can be reframed to better correspond to policy choices.

While the specific examples of beneficial reframing considered in this thesis are useful in themselves, a broader import of the thesis as a whole is the demonstration of the need for critical reflection within CEA with respect to model framing. This process of critical reflection should include a consideration of what exactly is decision problem at hand, whether this is adequately reflected by the model and are there any assumptions implicit in the model that compromise its fit to the policy question. Any failure to ask these questions may lead to an inadequately framed model, which in turn may lead to misguided policies, wasted resources and suboptimal health outcomes. Hopefully the work contained in this thesis will, in a small way, help avoid such outcomes.

REFERENCES

1. Westra TA, Parouty MBY, Wilschut JC, Boersma C, Postma MJ. RE Practical Implications of Differential Discounting of Costs and Health Effects in Cost-Effectiveness Analysis. Value in Health. 2012;14(8):1173-4.

2. O'Mahony JF, de Kok IMCM, van Rosmalen J, Habbema JDF, van Ballegooijen M. RE Practical Implications of Differential Discounting of Costs and Health Effects in Cost-Effectiveness Analysis. Value in Health. 2012;14(8):1174-5.

3. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of health care technologies. Health Economics. 2011;20(1):2-15.

4. van Rosmalen J, de Kok I, van Ballegooijen M. Cost-effectiveness of cervical cancer screening: cytology versus human papillomavirus DNA testing. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(6):699-709.

5. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of Human Papillomavirus vaccination in the Netherlands. Journal of the National Cancer Institute. 2009;101(15):1083-92.

6. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–1. Medical Decision Making. 2012;32(5):667-77.

7. NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Clinical Excellence, 2004.

8. NICE. Guide to the Methods of Technology Appraisal. London: National Institute for Health and Clinical Excellence, 2008.

9. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada, 3rd Edition. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2006.

10. HIQA. Guidelines for the Economic Evaluation of Health Technologies in Ireland. Dublin: Health Information and Quality Authority, 2010.

11. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M. Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. Medical Decision Making. 2012;32(5):678-89.

12. Torrance GW, Siegel JE, Luce BR. Framing and Designing the Cost-Effectiveness Analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 54-81.

13. Caro JJ, Möller J, Getsios D. Discrete Event Simulation: The Preferred Technique for Health Economic Evaluations? Value in Health. 2010;13(8):1056-60.

14. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. Medical Decision Making. 2012;32(5):722-32.

15. Craig BM, Busschbach JJV. Toward a more universal approach in health valuation. Health Economics. 2011;20(7):864-75.

APPENDIX OF OTHER WORK

Overview

This appendix contains work other than the six papers included in the body of the thesis. The purpose of its inclusion is to provide evidence of other work undertaken during the course of the thesis. This additional material includes a brief comment piece forthcoming in Value in Health regarding NICE's recent decision to adopt differential discounting in selected circumstances; a co-authored paper with colleagues from Erasmus University on a continuous time alternative to conventional discrete time Markov modelling with an application to a Markov model of Hepatitis B control, recently accepted for publication in Medical Decision Making; two letters regarding aspects of methods within CEA; and, seven posters presented during the course of the research for this thesis, including two considering cost-effectiveness acceptability curves as a measure of uncertainty in CEA.

NICE's Selective Application of Differential Discounting: Ambiguous, Inconsistent and Unjustified

James F O'Mahony, MA[1],
Mike Paulden, MSc, MA[2].

1. Department of Health Policy and Management, Trinity College Dublin, Ireland
2. Toronto Health Economics and Technology Assessment (THETA) Collaborative, University of Toronto, Toronto, Canada

Forthcoming in Value in Health

ABSTRACT

The National Institute for Health and Clinical Excellence (NICE) recently recommended differential discounting of costs and health effects in the economic appraisal of healthcare interventions in certain circumstances. The recommendation was published in an amendment to the guide to the methods of technology appraisal. It states that differential discounting should be applied where "treatment effects are both substantial in restoring health and sustained over a very long period (normally at least 30 years)". Renewed interest in differential discounting from NICE is welcome; however, the recommendation's selective application of differential discounting raises a number of concerns. The stated criteria for applying differential discounting are ambiguous. The rationale for selective application of differential discounting has not been articulated by NICE and is questionable. The selective application of differential discounting leads to several inconsistencies, the most concerning of which is the lower valuation of health gains for those with less than 30 years remaining life expectancy, which can be interpreted as age discrimination. Furthermore, the discount rates chosen by NICE do not appear to be informed by recent advances in the understanding of differential discounting. Finally, NICE's apparent motivation for recommending differential discounting was to ensure a favourable cost-effectiveness ratio for a paediatric oncology drug. While flexibility may be appropriate to allow some interventions that exceed conventional cost-effectiveness thresholds to be adopted, the selective adjustment of appraisal methods is problematic and without justification.

Background

In July 2011, NICE published an amendment to their methods guidance for the economic evaluation of health technologies regarding the discounting of costs and health effects [1]. Since 2004 NICE has recommended equal discounting of costs and effects at a rate of 3.5% per annum. However, the recent amendment states that costs and health effects should now be differentially discounted at 3.5% and 1.5% per annum respectively in specific cases where "treatment effects are both substantial in restoring health and sustained over a very long period (normally at least 30 years)". Such differential discounting will generally result in the health technology in question having a more favourable cost-effectiveness estimate than under equal discounting, in turn strengthening the case for its adoption.

NICE's amendment was made following an appraisal committee's consideration of mifamurtide, a drug indicated for osteosarcoma, a rare disease which principally afflicts children and young adults. An article published on NICE's website explained the appraisal committee's deliberation over the drug's cost-effectiveness and the decision to apply differential discounting [2]. The article notes that, under NICE's standard 3.5% discount rate, mifamurtide's incremental cost-effectiveness ratio (ICER) was estimated to be £57,000 per quality-adjusted life year (QALY), considerably higher than NICE's usual threshold range of £20,000 to £30,000 per QALY. The appraisal committee noted that applying differential discounting at 3.5% & 1.5% for costs and health effects respectively, reduced the estimated ICER to a more favourable £36,000 per QALY. While this ICER remains above the threshold range, it is broadly similar to those of other interventions approved by NICE given special considerations such as disease severity and childhood disease [3], and the NICE appraisal committee recommended the adoption of mifamurtide.

In this article, we do not wish to address the merits of recommending mifamurtide. We understand and respect the appraisal committee's decision. However, the role of NICE's selective application of differential discounting in the adoption decision and the subsequent amendment to NICE's methods guidance is concerning and deserves comment.

We describe how NICE's decision to recommend the selective use of differential discounting raises a number of methodological difficulties and inconsistencies potentially undermine the scientific rigour of NICE's economic evaluation process. The language of NICE's amendment is ambiguous, raising further difficulties of interpretation. These ambiguities, difficulties, and inconsistencies are summarised below.

Sensitivity analysis or reference case?

The first ambiguity is whether the differential rates of 3.5% & 1.5% are the "reference case" rates for the special cases according to the given criteria or if they are required only as part of a sensitivity analysis. The amendment notes that the exiting NICE guidance recommends conducting a sensitivity analysis with rates between 0% & 6%, including differential discounting. The amendment states that if a sensitivity analysis is undertaken, then rates of 3.5% & 1.5% should be applied rather than the standard reference case rates of 3.5% equal discounting. The juxtaposition with the standard rate of 3.5% strongly implies the 3.5% & 1.5% rates are the reference case rates for the special cases. Furthermore, the NICE commentary on the use of differential discounting in the appraisal of mifamurtide also supports this interpretation. However, the reference to sensitivity analysis makes it unclear if this interpretation is correct.

Eligibility for differential discounting

A further ambiguity relates to which cases are eligible for differential discounting. The guidance recommends differential discounting where health gains are sustained. What is meant by sustained is not clear. It could preclude interventions that yield a long-term health gain, but only with the maintained course of an intervention rather than once-off treatment, such as HIV anti-retroviral therapy. The amendment also states that the intervention must be substantial in restoring health. Presumably the restorative criterion is intended to preclude preventative interventions. However, it is unclear if this requires patients to have first suffered a health decrement before the intervention alleviates this burden, or whether it is sufficient for the intervention to halt disease progression from a healthy state to lower health state. The former would seem a strict interpretation of restoring health, while the latter would seem to apply to mifamurtide. Finally, it is also unclear how great the health effects need be

to qualify as substantial. For example, there could be differences of interpretation of substantial regarding relative or absolute improvements in health and whether gains need be substantial at the individual or population level.

Inconsistencies resulting from the eligibility criteria

NICE's recommendation that differential discounting be selectively applied in some case but not others gives rise to apparent inconsistencies, whereby interventions of similar characteristics are subject to different discounting assumptions, potentially leading to large differences in cost-effectiveness. Four apparent inconsistencies are as follows:

1. If we assume the restorative criterion precludes preventative care, then we can consider two strategies to control the same disease, one preventative and one curative, such as vaccination against the Human Papillomavirus and treatment for cervical cancer. Assuming both interventions satisfy the other criteria, then although they potentially achieve the same outcome their effects will be discounted differently. In this case, vaccination would be disadvantaged relative to treatment in terms of cost-effectiveness.
2. Similarly, if the condition of sustained health gain precludes interventions that require maintained therapy rather than a one-off intervention, then a maintained intervention such as anti-retroviral therapy will be subject to equal discounting, while a drug like mifamurtide enjoys the benefit of differential discounting.
3. Consider two interventions that yield the same aggregate QALY gain, one achieving this by bringing about a small QALY gain per individual in a large patient population, while another achieves a large QALY gain per individual in a small patient population. If the substantial criterion precludes the application of differential discounting in interventions with small health gains per patient, then despite having similar aggregate QALY gains, the two interventions may be subject to different discounting rates.
4. Finally, consider two interventions, the first yielding benefits for 29 years, the second yielding identical benefits for 30 years. Assume both meet the amendment's other criteria. Despite the fact that both interventions produce the same benefit for 29 of the 30 years, the second will qualify for differential discounting, while the first will not. This means that the initial

29 years of identical benefits will be valued differently by NICE's new discounting scheme, simply because the second intervention achieved one more year of benefits.

Discrimination on the basis of life expectancy

The final inconsistency listed above merits particular consideration. The criterion that health gains must be sustained for 30 years or more creates scope for arbitrary discrimination solely on the basis of life expectancy. It means an individual with a remaining life expectancy of 30 years or more could be eligible for differential discounting, whereas a marginally older individual with less remaining life expectancy would be subject less favourable equal discounting. Consequently, there could be a large difference in the intervention's cost-effectiveness estimates between these two similar individuals, potentially leaving the younger eligible for treatment, but not the older. Such arbitrary discrimination against those with shorter life expectancy is unjustified. Furthermore, since individuals with shorter life expectancy are often (but not always) older than those with longer life expectancy, the revised discounting guidance potentially exposes NICE to accusations of ageism.

NICE has previously been accused of both ageism and discrimination against those with short life expectancy, most notably by Harris [4]. However, this criticism has been countered by pointing out that NICE does not value health gains in older patients or those with short life expectancy any less than equivalent health gains in other patients [5]. Unfortunately, with NICE's selective application of differential discounting this defence may no longer stand in all cases.

Theoretical issues and the choice of discount rates

NICE's amendment is also problematic regarding the discount rates it recommends. The amendment makes no reference to recent theoretical work on the appropriate discount rates by Claxton et al. and Paulden and Claxton [10, 11]. Claxton et al. demonstrate that, in a publicly funded health system subject to a fixed budget constraint, the differential between the discount rates on costs and effects should approximate the real annual growth rate of the cost-effectiveness threshold, with positive threshold growth corresponding to a lower discount rate for effects.

However, NICE's stated cost-effectiveness threshold range has not changed in nominal terms over recent years, implying a fall in real terms [12]. Consequently, it is difficult to see how the 2% differential used by NICE is justified. Furthermore, Paulden and Claxton show that the discount rate on costs should approximate the real rate of return on government bonds. The real rate of return on UK government bonds is currently far below the 3.5% used by NICE and has been for some time. Consequently, there is a notable irony that if NICE had set the discount rates for costs and effects equal to each other at current real government bond yields then mifamurtide might have been found to be cost-effective and there would have been no need to adopt theoretically unjustified differential discounting.

Absence of rationale & NICE's Citizen's Council deliberation of differential discounting

An overarching problem with the amendment is the lack of rationale for the selective application of differential discounting. The amendment gives no justification for eligibility criteria it sets and there is no obvious reason why they should apply. While NICE give no rationale for selective differential discounting some insight is offered by a recent consultation of NICE's Citizen Council of lay people. In November 2011 NICE asked its Citizen's Council to consider the issue of discounting. A final report of this consultation was published in August 2012 [6].

As part of the consultation, NICE asked the council to consider differential discounting in the context of a hypothetical example of a highly effective (curative in most cases), but very costly orphan drug used to control a childhood disease that is usually fatal if unchecked. The ICER of the drug under 3.5% equal discounting was given as £57,000/QALY and £24,000/QALY under differential discounting at 3.5% & 1.5%. This example clearly and apparently deliberately resembles mifamurtide. The Citizen's Council expressed concern regarding the implications of discounting at a common rate of 3.5% per annum and a majority of the council said they would adopt the hypothetical drug. The council agreed that differential discounting should be applied in cases of highly effective or curative interventions, interventions with health effects over a long period of time and (less unanimously) interventions for children.

Despite the Citizen's Council apparent approval of differential discounting in the mifamurtide-like example, it is doubtful if this lends credibility to NICE's selective application of differential discounting. The example presented to the Citizen's Council conflates discounting with the issues of curative care, childhood disease and disease severity and rarity. Empirical evidence indicates that people have a stronger preference for health gains in children over adults, for health gains in those with severe illnesses over less serious conditions and for treatment over prevention [7, 8]. While there is no empirical evidence of individuals having stronger preferences for treating rare diseases, there is a an apparent degree of policy support for treating rare diseases preferentially, as evidenced by financial incentives for research on such conditions in the US and EU [9]. Consequently, the Citizen's Council approval of differential discounting in the mifamurtide-like example may simply be an expression of a greater willingness to adopt such an intervention, rather than an endorsement of selective differential discounting per se.

Moving forwards

While NICE's selective adoption of differential discounting is an easy target for criticism, we do have sympathy with those deciding whether or not to approve mifamurtide. Rejecting mifamurtide would likely prompt strong criticism of NICE by the popular press. Faced with public pressure and an apparent societal preference to prioritise curative care for severe, rare diseases in children a degree of flexibility was not unwise. However, we feel such flexibility should be exercised in a transparent and scientifically rigorous manner. If NICE wishes to incorporate concern for particular patients or diseases into its cost-effectiveness analyses, then this should be done so explicitly through means such as a higher threshold or QALY weights in clearly defined eligible cases. This approach already has precedence with NICE's consideration of "end of life care" and in other decisions where interventions with ICERs over the threshold range have been adopted for explicit reasons [3, 13].

In summary, NICE's recently revised discounting guidance gives rise to numerous ambiguities, difficulties, and inconsistencies, which have the potential to undermine the scientific rigour of NICE's economic evaluation process. In particular, NICE's amendment creates scope for arbitrary discrimination on the basis of life

expectancy. Flexibility in decision making will always be necessary; however, the decision to achieve this flexibility in the case of mifamurtide by selectively applying differential discounting has created scope for confusion and inconsistency, with potentially unforeseen consequences. Furthermore, the discount rates chosen are not supported by our current understanding of differential discounting. We hope that NICE will consider these issues and refer to the body of theoretical research on discounting when revising their discounting methodology in the future.

REFERENCES

1. NICE, Discounting of Health Benefits in Special Circumstances. 2011, National Institute of Health and Clinical Excellence.

2. NICE. How should NICE assess future costs and health benefits? 2011 [cited 19-6-2012]; Available from: http://www.nice.org.uk/newsroom/features/HowShouldNICEAssessFutureCostsAndHealthBenefits.jsp.

3. Rawlins, M., D. Barnett, and A. Stevens, Pharmacoeconomics: NICE's approach to decision-making. British Journal of Clinical Pharmacology, 2010. 70(3): p. 346-349.

4. Harris, J., QALYfying the value of life. Journal of Medical Ethics, 1987. 13(3): 117-123.

5. Paulden, M. and A.J. Culyer, Does cost-effectiveness analysis discriminate against patients with short life expectancy? Matters of logic and matters of context, in Theta Collaborative Working Paper Series. 2010, University of Toronto: Toronto.

6. NICE, How should NICE assess future costs and health benefits. 2012, National Institute of Health and Clinical Excellence: London.

7. Dolan, P., et al., QALY maximisation and people's preferences: a methodological review of the literature. Health Economics, 2005. 14(2): 197-208.

8. Corso, P.S., et al., Assessing Preferences for Prevention versus Treatment Using Willingness to Pay. Medical Decision Making, 2002. 22(S1): s92-s101.

9. Mentzakis, E., P. Stefanowska, and J. Hurley, A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. Health Economics, Policy and Law, 2011. 6(03): 405-433.

10. Claxton, K., et al., Discounting and decision making in the economic evaluation of health care technologies. Health Economics, 2011. 20(1):2-15.

11. Paulden, M. and K. Claxton, Budget allocation and the revealed social rate of time preference for health. Health Economics, 2011. 21(5): 612-618.

12. McCabe, C., K. Claxton, and A.J. Culyer, The NICE Cost-Effectiveness Threshold: What it is and What that Means. PharmacoEconomics, 2008. 26(9): 733-744.

13. NICE, Supplementary Advice to the Appraisals Committees: Appraising life-extending, end of life treatments. 2009.

A Mathematical Approach for Evaluating Markov Models in Continuous Time without Discrete-Event Simulation

Joost van Rosmalen, PhD[1],
Mehlika Toy, PhD[1,2],
James F O'Mahony, MA[1,3].

1. Department of Public Health, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
2. Takemi Fellow, Department of Global Health and Population, Harvard School of Public Health. Boston, USA
3. Department of Health Policy and Management, Trinity College Dublin, Dublin, Ireland

Medical Decision Making 33(3).

ABSTRACT

Background: Markov models are a simple and powerful tool for analyzing the health and economic effects of healthcare interventions. These models are usually evaluated in discrete time using cohort analysis. The use of discrete time assumes that changes in health states occur only at the end of a cycle. This assumption can yield biased cost-effectiveness estimates for Markov models with long cycle periods and if no half-cycle correction is made.

Objective: To evaluate Markov models in continuous time without using discrete-event simulation.

Methods: We use results from stochastic process theory and control theory to obtain a mathematical solution for the expected time spent in each state in a continuous-time Markov model (CTMM). We then show how this solution can account for age-dependent transition rates, discounting of costs and health effects, and transition rates that depend on the time spent in a state. To illustrate these methods, we evaluate a discrete-time Markov model (DTMM) of the cost-effectiveness of antiviral therapies for chronic hepatitis B using different cycle lengths, and by converting it into a CTMM.

Results: The cost-effectiveness results show pronounced differences between a DTMM with an annual cycle and the CTMM (e.g. incremental cost-effectiveness ratios of 4,000 instead of 4,800 euros per quality-adjusted lifeyear). The differences between the CTMM and DTMMs with cycle lengths of 1 month and 0.001 years are small and negligible, respectively. The CTMM required approximately 3 times more computation time than a DTMM with an annual cycle, but less computation time than DTMMs with a monthly cycle or a cycle of 0.001 years.

Conclusion: Evaluating Markov models in continuous time is a feasible alternative to cohort analysis and can offer several theoretical and practical advantages.

INTRODUCTION

Markov models are a simple and powerful tool for analyzing the health and economic effects of healthcare interventions [1-3]. In their seminal paper, Beck and Pauker [1] identified three methods to evaluate Markov models: a) cohort analysis, b) Monte Carlo simulation or microsimulation, and c) a fundamental matrix solution. Cohort analysis is a relatively simple technique that can often be done with a spreadsheet program [4]. Microsimulation is a flexible and increasingly popular alternative to cohort analysis, as this method can account for a patient's history of prior events and can also be applied to disease processes that are too complex to evaluate using Markov models [5]. Finally, the fundamental matrix solution uses matrix algebra to obtain a quick and exact solution for the time spent in each state in Markov chain models.

Markov models can be evaluated in either discrete time or continuous time. In discrete-time Markov models (DTMMs), the time period over which an intervention's effects are assessed is divided into (usually) fixed time intervals or cycles. Patients can only undergo a state transition at the end of a cycle period, subject to the specified transition probabilities. In a continuous-time Markov model (CTMM), transitions can occur at any point in time. A CTMM can be seen as a DTMM with an infinitesimally short cycle. DTMMs are far more common than CTMMs in the cost-effectiveness analysis (CEA) of healthcare interventions [6].

DTMMs use discrete time to approximate a continuous-time disease process. In reality, changes in health states do not occur only at the end of an artificially specified cycle period. By summing state membership at the beginning or the end of each cycle, DTMMs lead to over- and underestimation of the time spent in each state, respectively. These biases can be made considerably smaller by employing the half-cycle correction (i.e. adding a half-cycle's worth of incremental utility to the cumulative total for each health state [7]) or averaging state membership at the beginning and the end of the cycle [8]. Alternatively, these biases can be attenuated by using short cycle lengths [9, 10]. However, these methods yield only approximate solutions and have the other disadvantages that short cycles lead to

long computation times and that the half-cycle correction is incompatible with discounting [8]. A real solution to the problem that state membership is not known between cycles would be to use CTMMs instead of DTMMs, as they provide exact results for the time spent in a state.

In addition to the discrete or continuous time distinction, models can also be categorized as deterministic (i.e. models that can be evaluated analytically) or stochastic (i.e. models that are simulation-based) [11]. Cohort analysis of DTMMs and Beck & Pauker's fundamental matrix solution are deterministic, whereas microsimulation models are stochastic. Deterministic models yield exact solutions, whereas stochastic models suffer from first-order simulation error.

This paper describes a continuous-time approach to Markov models in combination with a deterministic solution. Deterministic CTMMs avoid the problems of DTMMs and benefit from the advantages over simulation-based approaches. Methods to evaluate CTMMs as deterministic models are not known in the medical decision making literature. Rather, CTMMs have usually been analyzed as stochastic models, using discrete-event simulation (DES) [12]. However, DES models require much more time to develop and evaluate than Markov models, and DES typically requires dedicated DES software [13].

The purpose of this paper is to present methods to evaluate CTMMs in a deterministic way (i.e. without using DES) and to show how CTMMs can be used as an alternative to DTMMs. These methods for evaluating CTMMs are based on mathematical results from stochastic process theory and control theory. The mathematical approach used in this paper is not novel; our contribution is to show how this approach can be applied to Markov models used in CEA. An outline of this paper is as follows. First, an overview of the differences and similarities between CTMMs and DTMMs is presented. It is then shown how CTMMs can be evaluated using a deterministic solution, and how this solution can account for age-dependent transition rates and for discounting of costs and health effects. Furthermore, it is shown how tunnel states, which are often used in DTMMs [2], can be extended to CTMMs, to model transition probabilities that depend on the time spent in a state. An example of a Markov model of antiviral therapies for Hepatitis B is used to

illustrate CTMMs. Finally, the practical applicability of CTMMs in medical decision making is discussed and compared with other modeling approaches. A proof of the deterministic solution for CTMMs and programming code to apply this solution are given in the appendices.

Continuous-Time versus Discrete-Time Markov Models

Previous Literature on Continuous-Time Modeling

Markov models used in CEAs of healthcare interventions are almost always based on a discrete-time approach, with few exceptions [14]. However, several other techniques for medical decision analysis assume that disease progression occurs in continuous time. For example, Hazen [15] proposed continuous-time decision trees. In addition, DES models often assume a set of mutually exclusive health states and disease progression in continuous time [5, 16]. If the sojourn time distributions also do not depend on which states were visited previously, these DES models can be seen as continuous-time Markov (if all sojourn time distributions are exponential) or semi-Markov (if some sojourn time distributions are not exponential) models.

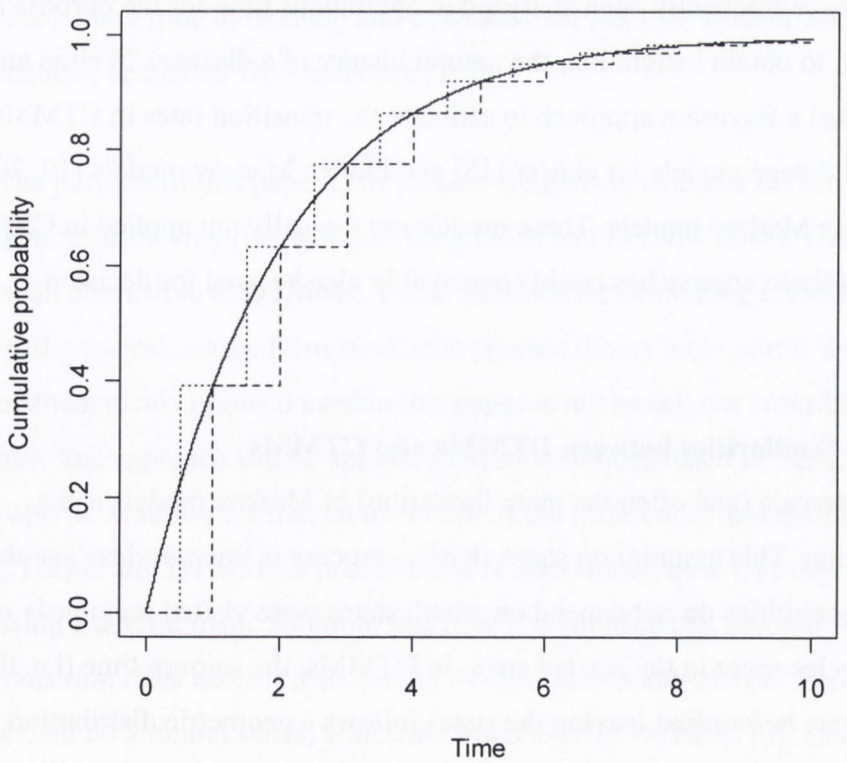
Markov models have frequently been analyzed in continuous time for the purpose of data analysis (e.g. to obtain insight into the natural history of a disease). Welton and Ades [17] proposed a Bayesian approach to estimate the transition rates in CTMMs. Furthermore, multistage models for cancer [18] and hidden Markov models [19, 20] can also be seen as Markov models. These models are typically not applied in CEA, although some of these approaches could conceivably also be used for decision analysis.

Differences and Similarities between DTMMs and CTMMs

The main characteristic (and often the main limitation) of Markov models is the Markov assumption. This assumption states that the process is 'memoryless', so that the transition probabilities do not depend on which states were visited previously or the number of cycles spent in the current state. In DTMMs, the sojourn time (i.e. the time spent in a state before first leaving the state) follows a geometric distribution, with a mean duration of $1/p$ cycles, where p is the per-cycle transition probability. In CTMMs, the Markov assumption results in an exponential distribution for the

sojourn time, with a mean duration of $1/r$, where r is the transition rate out of the current state. Figure 1 shows the cumulative distribution function of an exponential distribution with $r = 0.5$ and a geometric distribution with $p = 1 - \exp(-0.5)$, both with and without half-cycle correction. This figure shows that a CTMM (with an exponential distribution) yields a lower expected time spent in the state than a DTMM without half-cycle correction (with a geometric distribution). This bias in the expected time spent in a state is considerably reduced by applying a half-cycle correction (see Figure 1). Alternatively, this bias can be reduced by choosing a shorter cycle period in the DTMM, which leads to a lower probability of transition during a cycle, resulting in smaller differences between the geometric and exponential distributions.

Figure 1: Cumulative probability of having left the current state for a DTMM without half-cycle correction (dashed line), a DTMM with a half-cycle correction (dotted line), and a CTMM (solid line), as a function of the time spent in the state.



DTMMs are specified using per-cycle transition probabilities, whereas CTMMs are specified using transition rates. The transition rates describe the relative speed with which patients move from one state to another. A transition rate from state i to state j, with $i \neq j$, can be described mathematically as the limit

$$q_{ij} = \lim_{t \to 0} \frac{p_{ij}^{(t)}}{t},$$

where q_{ij} is the transition rate from state i to state j, $p_{ij}^{(t)}$ is the transition probability from state i to state j, and t is the cycle period. In general, a transition probability can be transformed into a transition rate according to

$$q_{ij} \approx -\frac{\ln(1 - p_{ij}^{(t)})}{t}, \qquad (1)$$

see Sonnenberg and Beck [2]; however, this approximation is accurate only if $p_{ij}^{(t)}$ is small. Miller and Homan [21] provide further discussion of the differences between transition rates and transition probabilities.

While in general the results of DTMMs converge to those of CTMMs as the cycle length is reduced, not every DTMM has a continuous-time analogue. For example, a DTMM with positive per-cycle transition probabilities from state A to B and from state B to C, but with a zero transition probability between states A and C, cannot be represented as a CTMM. The reason being is that in a CTMM with positive transition rates from state A to B and from state B to C, there is always a positive probability that a patient undergoes both transitions (from states A to B to C) during any positive time interval. Thus, for some Markov models, the Markov assumption no longer holds with shorter cycle lengths or in continuous time. Whether a DTMM can be represented using an underlying CTMM is known as the embeddability problem [22, 23].

The cycle length in DTMMs should be chosen with care. Sonnenberg and Beck [2] argued that the cycle period should be set to a 'clinically meaningful time interval'. However, this criterion seems somewhat ambiguous, and in practice, the cycle period is usually set to a period of 1, 6, or 12 months. The cycle period should not be too long, as the assumption that transitions occur only at the end of a cycle

period can lead to a bias in the estimated amount of time spent in a state and to a bias in the estimated cost-effectiveness ratios. A very short cycle period will avoid these biases, but may be computationally intensive.

DTMMs can offer some flexibility for modeling disease processes that do not satisfy the Markov assumption. For example, DTMMs can include temporary states (i.e. states from which the patient leaves during a cycle with probability 1) to model health states that have a fixed duration [2], which is not possible in CTMMs. DTMMs can also include tunnel states, i.e. a number of temporary states such that each state has a transition only to the next state [2]. The use of temporary states and tunnel states enables a modeler to account for transition probabilities that depend on the time spent in a health state, but can also lead to a proliferation of the number of states. In the next section, we show how tunnel states and similar approaches can be used in CTMMs.

One remaining obstacle for the use of CTMMs is that easy-to-use solution strategies (i.e. techniques for determining the expected amount of time spent in each state) have not been presented in the medical decision making literature. In the next section, we show how the required results can be calculated mathematically without using DES.

Mathematical Results for CTMMs

Consider an example of a simple discrete-time Markov chain model of a hypothetical disease with 3 health states ('well', 'ill', and 'dead'). The annual transition probabilities can be gathered in a matrix **P**, where element p_{ij} denotes the transition probability from state i to state j. The matrix **P** for an assumed set of transition probabilities is given by

$$\begin{array}{c} \\ \text{well} \\ \text{ill} \\ \text{dead} \end{array} \begin{array}{c} \begin{array}{ccc} \text{well} & \text{ill} & \text{dead} \end{array} \\ \begin{pmatrix} 0.93 & 0.05 & 0.02 \\ 0.3 & 0.5 & 0.2 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

The matrix of transition probabilities **P** can be used to compute the matrix of transition rates **Q**. For each pair of distinct states, the transition probabilities can be converted into transition rates (e.g. using equation 1), which yields the off-diagonal elements of **Q**. The diagonal elements of **Q** must contain the transition rates out of each state; these diagonal elements are calculated as $q_{ii} = -\sum_{j \neq i} q_{ij}$, so that each row of **Q** has sum 0. For example, the matrix **Q** corresponding to the probabilities assumed above, with the off-diagonal elements calculated using equation 1[1], is

$$\begin{array}{c} \\ \text{well} \\ \text{ill} \\ \text{dead} \end{array} \begin{array}{c} \begin{array}{ccc} \text{well} & \text{ill} & \text{dead} \end{array} \\ \begin{pmatrix} -0.071 & 0.051 & 0.020 \\ 0.357 & -0.580 & 0.223 \\ 0.000 & 0.000 & 0.000 \end{pmatrix} \end{array};$$

The mean sojourn time of state i is given by $-1/q_{ii}$, and the probability that a patient in state i first moves to state j is $-q_{ij}/q_{ii}$. The matrix of transition rates **Q** can thus be obtained using the mean duration of each state and, for each state, the relative probability distribution of the subsequent state.

In CEAs, the required results (e.g. expected discounted costs, health effects, and cost-effectiveness ratios) are typically derived from the expected distribution of the

[1] Please note that this univariate conversion of probabilities into rates does not account for the possibility of multiple transitions during a year, so that it can produce small changes in the implied annual transition probabilities and incorrect results in some applications.

patients over the states and the expected discounted time spent in a state. The calculation of the expected discounted time spent in each state is needed to determine the quality-adjusted life years (QALYs) gained or lost by an intervention.

Stochastic process theory [24] can be used to obtain the results mentioned above for CTMMs. We first consider the case of a continuous-time Markov chain model, in which the transition rates do not depend on age. Let $\pi_{i,t}$ be the probability that a patient is in state i at time t, let K be the number of Markov states, and let $\boldsymbol{\pi}_t = (\pi_{1,t} \quad \pi_{2,t} \quad \cdots \quad \pi_{K,t})$ be a row vector with the probability distribution of the Markov chain at time t. The state probability distribution $\boldsymbol{\pi}_t$ must satisfy the Kolmogorov forward equations [24], which are given by

$$\frac{\partial \boldsymbol{\pi}_t}{\partial t} = \boldsymbol{\pi}_t \mathbf{Q}.$$

The solution to these differential equations yields the state distribution of the Markov process at time t, and is given by

$$\boldsymbol{\pi}_t = \boldsymbol{\pi}_0 \exp(t\mathbf{Q}) = \boldsymbol{\pi}_0 \left(\sum_{i=0}^{\infty} \frac{t^i \mathbf{Q}^i}{i!} \right), \qquad (2)$$

where exp(x) denotes the matrix exponential function [24]. The matrix exponential function has been considered notoriously difficult to calculate numerically [25]. Fortunately, functions for numerically calculating the matrix exponential with a sufficient degree of accuracy are currently available in statistical software packages such as R and MATLAB. Closed-form mathematical solutions are sometimes also possible, if the Markov model has just a few states [17].

Based on equation 2, the expected discounted time spent in each state until time T is given by

$$\int_{t=0}^{T} e^{-rt} \boldsymbol{\pi}_0 \exp(t\mathbf{Q}) dt, \qquad (3)$$

where r is the continuous-time discount rate and e^{-rt} is a continuous-time discount factor [14]. The continuous-time discount rate is calculated from an annual discount rate according to $r = \ln(1 + DR/100)$, where DR is the annual discount rate

expressed as an percentage. For example, an annual discount rate of 3% corresponds to a continuous-time discount rate of $r = \ln(1+3/100) \approx 0.02956$.

There are several ways to evaluate the integral in equation 3, including numerical integration; however the integral can also be solved analytically, using mathematical results from control theory [26]. The latter approach gives the expected discounted time spent in state i is element i of

$$\int_{t=0}^{T} e^{-rt}\boldsymbol{\pi}_0 \exp(t\mathbf{Q})dt = \begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left(T\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}_0 \\ \mathbf{0}_{K\times 1} & \mathbf{Q}-r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K} \\ \mathbf{I}_K\end{bmatrix}, \qquad (4)$$

where $\mathbf{0}_{K\times 1}$ and $\mathbf{0}_{1\times K}$ are column and row vectors of zeros with length K, and $\mathbf{I}_K$ is the $K \times K$ identity matrix. A proof of equation 4 is provided in Appendix 1. Although equation 4 is a novel result, it is essentially an application of a mathematical result derived by Van Loan [26], and similar approaches have been used previously [27]. The right-hand side of equation 4 can be evaluated numerically using statistical software such as R, and an example in R code is included in Appendix 2.

If the costs and quality-of-life weights of the Markov model do not vary with age, the discounted costs and QALYs can be calculated from the results of equation 4. For example, the discounted number of QALYs can be determined by multiplying the expected discounted time spent in each state with the associated quality-of-life weights and summing the results over all states. The calculation of the discounted number of QALYs can be described mathematically as

$$\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left(T\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}_0 \\ \mathbf{0}_{K\times 1} & \mathbf{Q}-r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K} \\ \mathbf{I}_K\end{bmatrix}\mathbf{u}, \qquad (5)$$

where $\mathbf{u}$ is a column vector of quality-of-life weights. The parameter T denotes the time horizon of the Markov model. If the CTMM has an absorbing state that will eventually contain all patients (e.g. death), setting T to a sufficiently large value (e.g. $T = 120$ years) will ensure that the model is analyzed until all patients have entered the absorbing state.

Although the approach of equation 4 is not the only way to calculate the expected discounted time spent in a state, it is easy to program. For Markov models with

fewer than 300 states (i.e. the number of rows of the matrix of transition rates, including all tunnel states), a numerical evaluation of equation 4 is fast and accurate, with a computation time of less than 1 second. For large Markov models (e.g. with more than 1,000 states), alternative algorithms may be considered, based on numerically solving a system of differential equations, or a technique called uniformization [28, 29]. However, using these more elaborate methods is likely not necessary, as most Markov cost-effectiveness models have considerably fewer than 100 states.

Age-Dependent Transition Rates

The previous equations assumed a Markov chain model, in which the transition rates are constant. In most applications, transition rates are age-dependent, so these equations do not apply. However, the expected discounted time spent in a state can still be computed analytically if the transition rates change only a finite number of times, so that the transition rates are a piecewise-constant function of age. Similarly, the discounted QALYs can be computed analytically if the quality-of-life weights are a piecewise-constant function of age.

Assume that, between ages 0 and T, the transition rates and the quality-of-life weights change J times at ages t_j, $j = 1,\ldots,J$ and assume that $t_0 = 0$ and that t_{J+1} is infinity. Let $\mathbf{Q}_j$ be the matrix of the transition rates between ages t_j and t_{j+1}, and let $\mathbf{u}_j$ be the column vector of quality-of-life weights between ages t_j and t_{j+1}. The expected state distribution at age T, with $T \geq t_J$, is

$$\boldsymbol{\pi}_T = \boldsymbol{\pi}_0 \left(\prod_{j=0}^{J-1} \exp\left((t_{j+1} - t_j)\mathbf{Q}_j\right) \right) \exp\left((T - t_J)\mathbf{Q}_J\right), \qquad (6)$$

based on an application of equation 2 for each time period with constant transition rates. The expected discounted time spent in each state between ages 0 and T, with $T \geq t_J$, can be computed as

$$\sum_{j=0}^{J-1}\left(\exp(-rt_j)\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left((t_{j+1}-t_j)\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}^{(j)}\\ \mathbf{0}_{K\times 1} & \mathbf{Q}_j - r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K}\\ \mathbf{I}_K\end{bmatrix}\right)+$$
$$\exp(-rt_J)\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left((T-t_J)\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}^{(J)}\\ \mathbf{0}_{K\times 1} & \mathbf{Q}_J - r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K}\\ \mathbf{I}_K\end{bmatrix}, \quad (7)$$

where $\boldsymbol{\pi}^{(j)}$ represents the state distribution at age t_j and is calculated as $\boldsymbol{\pi}^{(j)} = \boldsymbol{\pi}^{(j-1)}\exp\left((t_j - t_{j-1})\mathbf{Q}_{j-1}\right)$. The discounted number of QALYs can be calculated as

$$\sum_{j=0}^{J-1}\left(\exp(-rt_j)\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left((t_{j+1}-t_j)\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}^{(j)}\\ \mathbf{0}_{K\times 1} & \mathbf{Q}_j - r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K}\\ \mathbf{I}_K\end{bmatrix}\mathbf{u}_j\right)+$$
$$\exp(-rt_J)\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{1\times K}\end{bmatrix}\exp\left((T-t_J)\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}^{(J)}\\ \mathbf{0}_{K\times 1} & \mathbf{Q}_J - r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K}\\ \mathbf{I}_K\end{bmatrix}\mathbf{u}_J.$$

(8)

Equations 7 and 8 can be derived by applying equation 4 for each time interval with constant transition rates and discounting all results to age 0. Appendix 2 contains an example of R code that implements equation 7.

In principle, equations 6, 7, and 8 only apply if the transition rates and the quality-of-life weights are piecewise-constant with respect to age. However, any function of age-specific transition rates can be approximated to any desired level of accuracy using piecewise-constant transition rates. In many published CEAs using Markov models, the transition probabilities change only a limited number of times. Thus, in practice, these equations can be applied in any CTMM. A disadvantage of the approach of equations 6, 7, and 8 is that the computation time increases with the number of age ranges used.

Transition Rates Depending on the Time Spent in a State

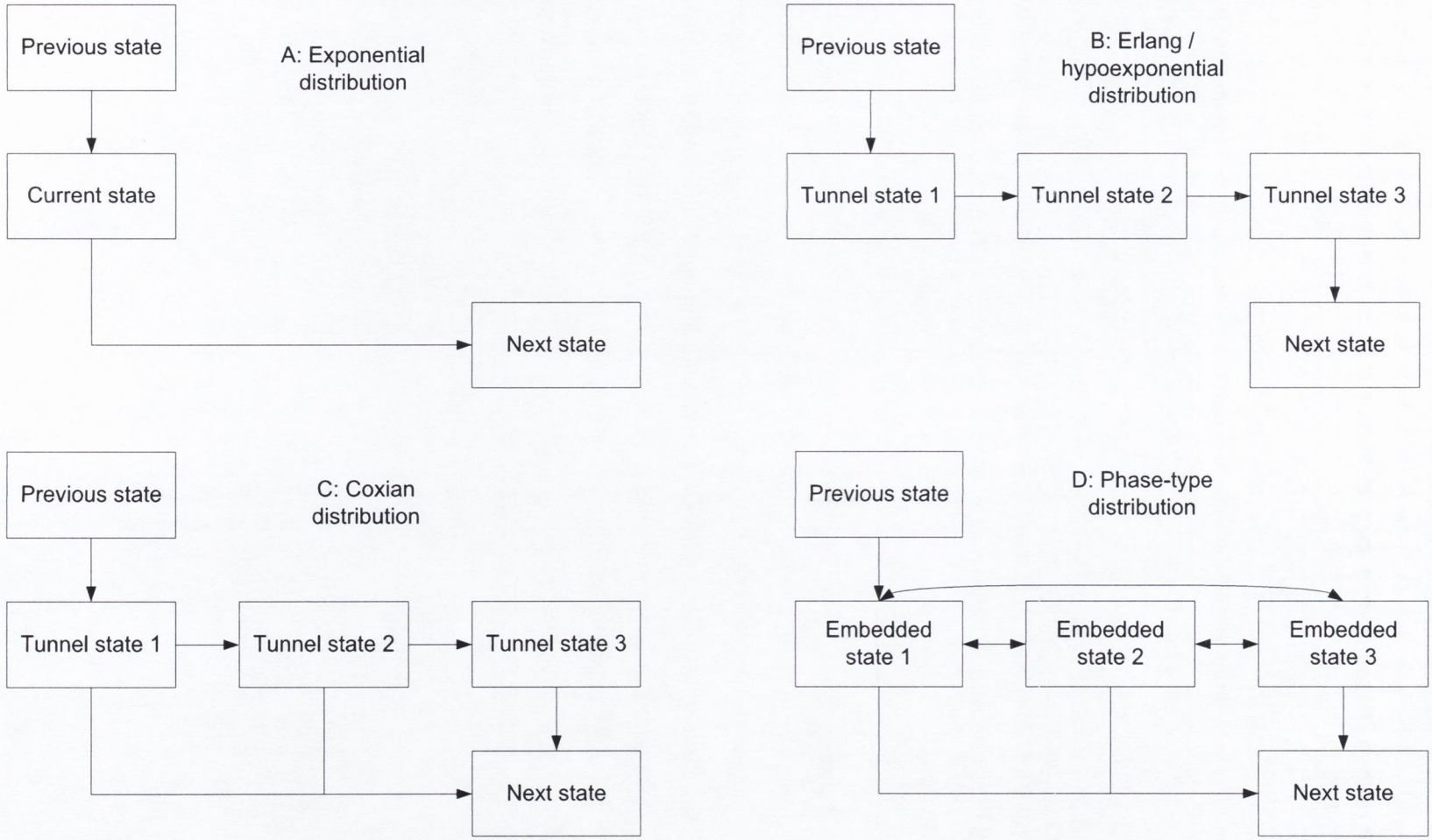
For CTMMs, the Markov property ensures that the sojourn time follows an exponential distribution. However, diseases do not always progress according to exponential duration distributions. The strong assumptions underlying the Markov property are thus an important limitation of Markov models. For DTMMs,

Sonnenberg and Beck [2] presented techniques (i.e. temporary states and tunnel states) to account for transition probabilities that depend on the time spent in a health state. A more elaborate approach, using 3-dimensional transition probability matrices, was proposed by Hawkins et al. [30]. Similar approaches can be used for CTMMs.

The general approach considered here is to represent a single health state using multiple states (called tunnel states or embedded states) in the Markov model. By including the possibility of transitions among the Markov states that represent a health state, it is possible to obtain a non-exponential sojourn time distribution for the set of Markov states that represent a health state, and thus to account for transition rates that depend on the time spent in a health state. Three possible strategies for modeling a non-exponential distribution for the sojourn time of a health state are as follows (see Figure 2 for a graphical illustration) [31].

1. Dividing a health state into a set of tunnel states that are all visited in a fixed sequence, with patients leaving the health state only from the last tunnel state (option B in Figure 2). This strategy leads to an Erlang distribution (if all tunnel states have the same transition rate) or a hypoexponential distribution [32] (if the tunnel states have different transition rates) for the time spent in the health state.
2. Dividing a health state into a set of tunnel states that are visited in a fixed sequence, but with a transition possible out of the health state from every tunnel state (option C in Figure 2). This strategy leads to a Coxian distribution [32] for the time spent in the health state.
3. Dividing a health state into a set of embedded states with transitions possible between each pair of health states (option D in Figure 2). This strategy leads to a phase-type distribution [33] for the time spent in the health state.

Figure 2: Representation of 4 types of sojourn time distributions that can be obtained in CTMMs by dividing a health state into multiple Markov states (i.e. tunnel states or embedded states).



Erlang distributions are useful for modeling disease processes with a gradually increasing hazard, such as the progression of preclinical cancer; however Erlang and hypoexponential distributions can approximate only a limited set of sojourn time distributions. Coxian and phase-type probability distributions can approximate any continuous sojourn time distribution (i.e. a continuous nonnegative distribution) with any desired degree of accuracy [34, 35]. Consequently, Coxian and phase-type distributions can be used to model any semi-Markov process as a Markov process. Because the Coxian distribution requires fewer parameters than the phase-type distribution, the Coxian distribution seems the most promising technique for approximating semi-Markov disease processes. Coxian and phase-type distributions have been used in medical applications, e.g. for modeling the length of stay in hospitals [34, 35].

Unfortunately, it is not easy to determine the parameters that provide the best approximation of a given sojourn time distribution using a Coxian or phase-type distribution, though several algorithms have been proposed in the literature [31, 36]. This approach can considerably increase the number of Markov states, especially if a close approximation is required. However, if the shape of a sojourn time distribution is estimated using observational data, a Coxian distribution with just a few tunnel states usually offers sufficient flexibility to account for transition rates that vary with the time spent in the health state.

Example: Cost-effectiveness of antiviral therapies for chronic hepatitis B

We demonstrate the use of CTMMs by creating the continuous-time analogue of an existing DTMM. The model used here is a CEA of antiviral therapies for chronic hepatitis B (CHB) in the Netherlands [37].

CHB can impose significant morbidity and mortality as a result of progression to liver disease. While vaccination against hepatitis B has significantly reduced the burden of disease, it only benefits those not yet infected [38]. Those already with active CHB must rely on antiviral therapy to control the disease. Therapy with antiviral agents suppresses virus replication, thus preventing progression to cirrhosis and hepatocellular carcinoma. Such therapies face the difficulties of drug resistance and high costs [37].

The original model was programmed in TreeAge. We replicated this model in R and created an analogous continuous-time model. The original CEA considered both hepatitis B e-antigen positive and negative patients; here we modelled only e-antigen negative patients. The e-antigen negative group can be further separated into those with and without cirrhosis. We modelled these subgroups separately from age 50 to age 95. The proportion of individuals alive at age 95 is assumed negligible.

The original model has been described elsewhere [38]. The transition probability matrix for the model's ten states at age 50 in one of the treatment arms is given in Figure 3. Patients in the CHB subgroup start in the CHB state and progress through the other states, as shown in Figure 3. The cirrhosis subgroup starts in the cirrhosis state.

We modelled the natural history of CHB and three antiviral monotherapies: Lamivudine (Lam), Tenofovir (TDF) and Entecavir (ETV). These antivirals are nucleoside analogue reverse transcriptase inhibitors that control Hepatitis B by interrupting viral replication. Lam is less expensive than the other therapies, but is less effective [37]. Patients receive supportive care in all arms of the model, including natural history. The transition probabilities and costs vary between treatment arms, whereas the utility weights for each state do not [39, 40].

Figure 3: Transition probability matrix of the original DTMM for Lam treatment in the CHB subgroup at age 50.

	SVR	CHB	CHB Resist	Cirr	Cirr Resist	DCC	HCC	LT	HBV Death	OC Death
SVR*	0.996	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
CHB	0.100	0.652	0.230	0.012	0.000	0.000	0.002	0.000	0.000	0.004
CHB with Resistance	0.045	0.000	0.907	0.000	0.030	0.000	0.015	0.000	0.000	0.004
Cirrhosis	0.000	0.000	0.000	0.710	0.230	0.016	0.017	0.000	0.024	0.004
Cirrhosis with Resistance	0.000	0.000	0.000	0.000	0.835	0.079	0.034	0.000	0.048	0.004
Decompensated Cirrhosis	0.000	0.000	0.000	0.000	0.000	0.704	0.000	0.032	0.260	0.004
Hepatocellular Carcinoma	0.000	0.000	0.000	0.000	0.000	0.000	0.633	0.012	0.351	0.004
Liver Transplant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.930	0.066	0.004
HBV Death†	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
Other Cause Death	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

*Sustained Virologic Response
†Hepatitis B Virus Death

The original model gathered transition rates from previous studies and meta-analyses and converted them to annual transition probabilities using the formula $P = 1\text{-}e^{-rt}$, where r is the transition rate and t is the cycle length. We used this formula to convert probabilities back to rates and to probabilities for different cycle lengths. This approach of converting probabilities to rates on an individual basis yields the transition rates from the studies that informed the transition probabilities. The original model used a cycle length of one year, without a half-cycle correction. We replicated the original model and then generated results with shorter cycles of one month and 0.001 years. We also ran the original model with a half-cycle correction as described by Naimark et al. [7].

The results for the CHB and cirrhosis subgroups are reported in Tables 1A and B, respectively, and are also shown in Figure 4. The tables report the net costs and effects of each treatment relative to the disease's natural history and the incremental cost-effectiveness ratios (ICERs) of TDF. The results of the DTMM with a 0.001 year cycle closely match those of the CTMM (the remaining differences disappear as the cycle length is further reduced – results not shown). For the CHB group, the difference in the ICER between the DTMM with an annual cycle and the CTMM is small. However, in the cirrhosis group the difference is pronounced, with the ICER being considerably larger in the CTMM. Furthermore, in the cirrhosis group, the results of a monthly cycle and of an annual cycle with a half-cycle correction still do not closely approximate those of the CTMM.

Tables 1A and B also report the computation time of each model using R (version 2.14.0) on a standard desktop computer. While the CTMM requires more time than the DTMM with an annual cycle, it requires less computation time than the DTMM with cycle lengths of one month or 0.001 years, being approximately 250 times faster than the latter.

Table 1A: Results of the CHB subgroup in the Markov model of antiviral therapies for Hepatitis B, evaluated using cohort analysis with various cycle lengths and the deterministic solution for CTMMs. All costs and effects are reported per patient.

		DTMM				CTMM
Cycle length, years		1	1	1/12	1/1000	N/A
Half cycle correction		No	Yes	No	No	N/A
Treatment	Outcome					
Lam	Costs, €	6,900	7,400	7,900	8,000	8,000
	QALYs	1.64	1.64	1.63	1.62	1.62
TDF	Costs, €	55,900	58,200	60,400	60,900	60,900
	QALYs	5.23	5.22	5.36	5.37	5.37
ETV	Costs, €	82,000	84,900	87,200	87,700	87,700
	QALYs	4.74	4.73	4.82	4.83	4.83
TDF	ICER	13,700	14,200	14,100	14,100	14,100
Run time, seconds*		0.04	0.04	0.40	31.37	0.12

*Average of 5 runs using R (version 2.14.0) on a desktop computer with an Intel Core2 Quad 3.0GHz processor

Table 1B: Results of the cirrhosis subgroup in the Markov model of antiviral therapies for Hepatitis B, evaluated using cohort analysis with various cycle lengths and the deterministic solution for CTMMs. All costs and effects are reported per patient.

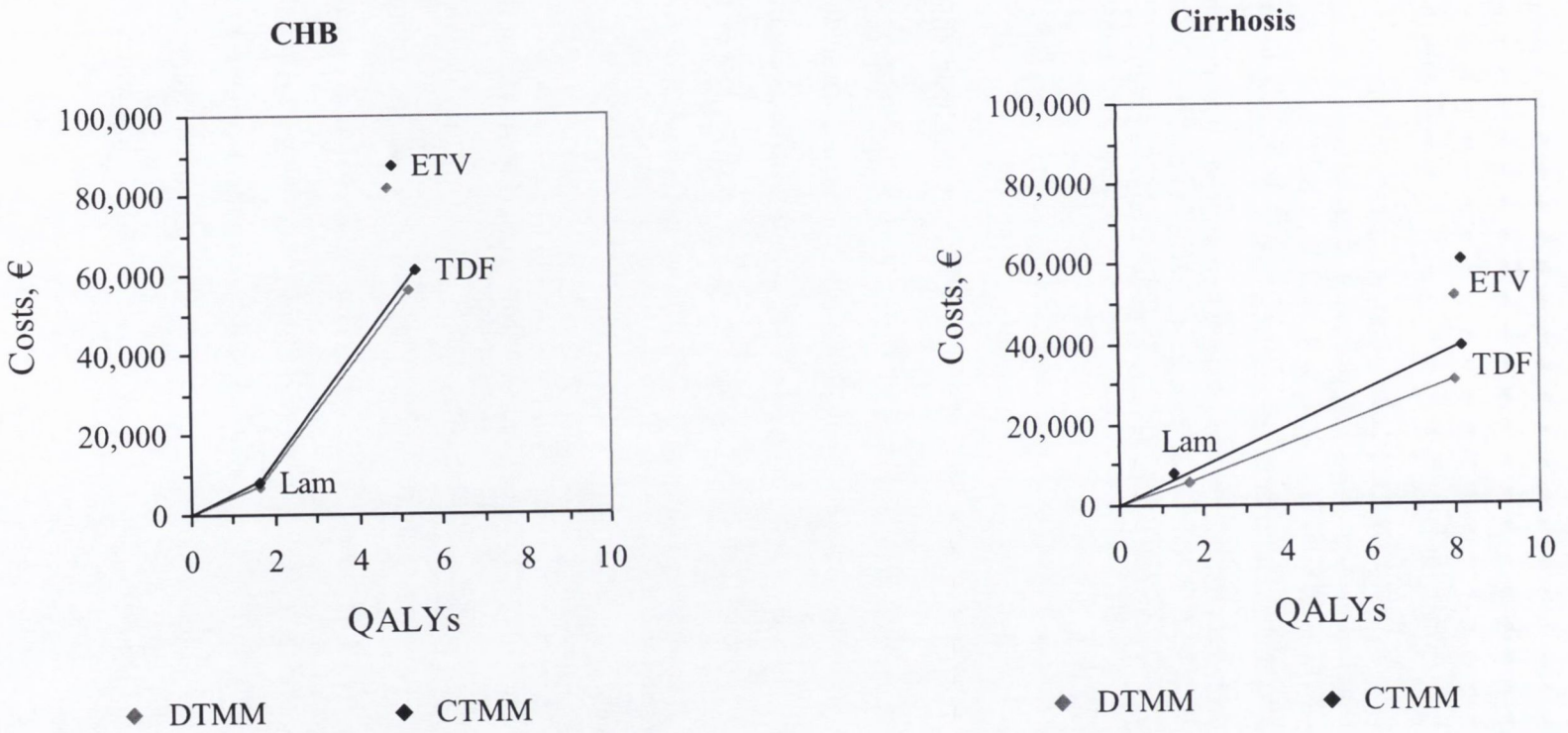
		DTMM				CTMM
Cycle length, years		1	1	1/12	1/1000	N/A
Half cycle correction		No	Yes	No	No	N/A
Treatment	Outcome					
Lam	Costs, €	5,500	6,000	7,800	8,000	8,000
	QALYs	1.70	1.70	1.34	1.31	1.31
TDF	Costs, €	30,800	33,100	38,300	39,000	39,000
	QALYs	7.98	7.97	8.14	8.15	8.15
ETV	Costs, €	51,700	54,600	60,100	60,900	60,900
	QALYs	7.98	7.97	8.14	8.15	8.15
TDF	ICER	4,000†	4,300†	4,700†	4,800‡	4,800‡
Run time, seconds*		0.03	0.04	0.39	32.33	0.13

*Average of 5 runs using R (version 2.14.0) on a desktop computer with an Intel Core2 Quad 3.0GHz processor

†TDF relative to Lam

‡ TDF relative to natural history

Figure 4: Cost-effectiveness planes depicting estimated costs and effects per patient of antiviral therapies for CHB (left panel) and cirrhosis (right panel) subgroups



The fairly large differences in cost-effectiveness results between the DTMM and the CTMM can partially be explained by the fact that the annual transition probability matrix implied by the CTMM does not exactly match that of the DTMM. This difference arises because a CTMM, unlike a DTMM, accounts for the possibility of multiple and competing events within one cycle of the DTMM [17, 41]. For example, the CTMM accounts for the possibility that, within a 1-year period, CHB patients may develop a) hepatocellular carcinoma directly or b) develop drug resistance and then hepatocellular carcinoma; a DTMM can only account for the first possibility.

DISCUSSION

Markov models have become a standard tool for CEAs of healthcare interventions, and cohort analysis in discrete time is the standard approach to evaluate these models. Although the Markov models used in medical decision making are seldom evaluated in continuous time, the mathematical results provided in this paper show that a continuous-time approach that does not require DES is feasible. The mathematics underlying our deterministic solution for CTMMs are not trivial; however, the solution itself is easy to apply, using widely available statistical software.

The principal advantage of deterministic CTMMs over DTMMs is that they provide an exact result for the time spent in each state and consequently they do not require a half-cycle correction or a short cycle length to achieve accurate results. Another advantage of CTMMs is that incidence rates observed in clinical or epidemiological studies can directly be used as model inputs, whereas DTMMs require the conversion of observed rates into transition probabilities. In CTMMs, it is relatively straightforward to estimate transition rates from observational data with different lengths of follow-up or if a patient's state is only known at fixed observation times [17, 42], which can require more effort using a DTMM [6, 43].

The advantages of CTMMs described above apply equally to DES models and the deterministic approach described here. However, a deterministic model also offers several further advantages over DES or microsimulation models. Deterministic CTMMs can be implemented in standard statistical packages such as R and are easier to create than DES models, which typically require specific DES software [13]. The deterministic solution for CTMMs typically requires less computation time than microsimulation. Therefore, CTMMs can be easier to use for computationally intensive techniques such as probabilistic sensitivity analyses and cost-effectiveness acceptability curves. In microsimulation models, these techniques often require nested simulations in which a large population is simulated for each set of simulated parameter estimates [43].

Another advantage of deterministic CTMMs over microsimulation models relates to optimization during model calibration, which is an important element of using either a Markov model or a microsimulation model. Model calibration is often done by searching for model parameters that ensure a good fit between the model outcomes and the observed data, sometimes using formal goodness-of-fit criteria and automated optimization algorithms [44]. In microsimulation models, the simulation error in the model outcomes can be a nuisance for model calibration. Due to the first-order simulation error in a microsimulation model, it is usually not possible to calculate the derivatives of the goodness-of-fit criterion, and these derivatives are necessary for most general-purpose optimization algorithms (i.e. the algorithms that are used for estimating the parameters of a logistic regression) [45]. Therefore, for microsimulation models, specialized optimization techniques are required instead [45-47]. For a deterministic CTMM, in which there is no first-order simulation error, and the main results are continuous functions of the model inputs, a goodness-of-fit criterion can usually be optimized using general-purpose optimization techniques (e.g. the optim() function in R).

A limitation of deterministic CTMMs is that these models require some programming and the use of the matrix exponential function. Consequently, CTMMs cannot easily be

applied using a spreadsheet program. Fortunately, deterministic CTMMs can be programmed in appropriate statistical software using only a few lines of code (see Appendix 2). Another limitation is that our mathematical approach for CTMMs may be considered less intuitive and less transparent than a spreadsheet-based technique that allows for a direct inspection of the intermediate results. A further limitation of CTMMs (and other types of Markov models) is that they cannot replace microsimulation models for some types of healthcare interventions. For example, in contrast to microsimulation models, Markov models do not allow researchers to incorporate interactions among individuals, which is required when modeling the transmission of infectious diseases.

Future research could focus on making CTMMs easier to apply in practice. An important element of this research could consist of making new methodologies available to estimate Markov transition rates from various types of observational data (e.g. Welton and Ades [17] presented a Bayesian approach). Such new methodologies could also include the estimation of the parameters of Coxian or phase-type sojourn time distributions. Finally, the applicability of CTMMs can be increased by including the mathematical results presented in this paper in future versions of healthcare decision analysis software.

CONCLUSION

Although DTMMs are currently far more popular than CTMMs, a continuous-time model can offer several theoretical and practical advantages over a discrete-time model. The use of CTMMs in CEA has hitherto been hampered by a lack of easy-to-use strategies for solving these models. The mathematical results presented in this paper make it feasible for modelers to apply CTMMs in CEAs.

APPENDIX 1: Proof of equation 4

The expected discounted time spent in each state is given by the integral

$\int_{t=0}^{T} e^{-rt}\boldsymbol{\pi}_0 \exp(t\mathbf{Q})dt$. For this integral, it holds that

$$\int_{t=0}^{T} e^{-rt}\boldsymbol{\pi}_0 \exp(t\mathbf{Q})dt = \int_{t=0}^{T} \boldsymbol{\pi}_0 \exp(-rt\mathbf{I}_k)\exp(t\mathbf{Q})dt = \int_{t=0}^{T} \boldsymbol{\pi}_0 \exp(t(\mathbf{Q}-r\mathbf{I}_k))dt,$$

where the last equality follows from the fact that $\exp(\mathbf{X})\exp(\mathbf{Y}) = \exp(\mathbf{X}+\mathbf{Y})$ if $\mathbf{X}$ and $\mathbf{Y}$ are commuting matrices. Van Loan [26] derived mathematical results for integrals involving matrix exponentials. He showed that if

$$\exp\left(T\begin{bmatrix}\mathbf{A}_1 & \mathbf{B}\\ \mathbf{0}_{n_2\times n_1} & \mathbf{A}_2\end{bmatrix}\right)=\begin{pmatrix}\mathbf{F}_1 & \mathbf{G}\\ \mathbf{0}_{n_2\times n_1} & \mathbf{F}_2\end{pmatrix}, \tag{9}$$

where $\mathbf{A}_1$, $\mathbf{A}_2$, and $\mathbf{B}$ are arbitrary matrices of sizes $n_1\times n_1$, $n_2\times n_2$, and $n_1\times n_2$, respectively, then $G=\int_0^T \exp((T-s)\mathbf{A}_1)\mathbf{B}\exp(s\mathbf{A}_2)ds$. Setting $\mathbf{A}_1=\mathbf{0}_{1\times 1}$, $\mathbf{B}=\boldsymbol{\pi}_0$, and $\mathbf{A}_2=\mathbf{Q}-r\mathbf{I}_k$ yields that

$G=\int_0^T \exp((T-s)\mathbf{0}_{1\times 1})\boldsymbol{\pi}_0\exp(s(\mathbf{Q}-r\mathbf{I}_k))ds=\int_0^T \boldsymbol{\pi}_0\exp(t(\mathbf{Q}-r\mathbf{I}_k))dt$, so that G is the value of the expected discounted time spent in each state. G can be computed by adding appropriate pre- and postmultiplications to equation 9, so that

$$G=\int_{t=0}^{T} e^{-rt}\boldsymbol{\pi}_0\exp(t\mathbf{Q})dt=\begin{bmatrix}\mathbf{I}_1 & \mathbf{0}_{K\times 1}\end{bmatrix}\exp\left(T\begin{bmatrix}\mathbf{0}_{1\times 1} & \boldsymbol{\pi}_0\\ \mathbf{0}_{1\times K} & \mathbf{Q}-r\mathbf{I}_K\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{1\times K}\\ \mathbf{I}_K\end{bmatrix},$$

which is the result in equation 4.

APPENDIX 2: R Script for Deterministic CTMMs

The following R script calculates the expected time spent in each state of a CTMM with age-dependent transition rates, based on equation 7. The 'Matrix' package is used for the calculation of the matrix exponential function.

```
library (Matrix)                          #Load the Matrix package
n <- 25                                   #Number of states
Q <- matrix(nrow=n,ncol=n)                #Matrix of transition rates
DR <- 3                                   #Annual discount percentage
r <- log(1+DR/100)                        #Continuous-time discount factor
T <- 50                                   #Time horizon (in years)
I <- diag(n)                              #Identity matrix
ZeroVector <- t(rep(0,n))                 #Row vector of zeros
InitStateDistr <- runif(n)                #Random initial state distribution
InitStateDistr <- t(InitStateDistr/sum(InitStateDistr))
TimeSpentInState <- t(rep(0,n))           #Initialize the discounted time spent in
each state
Interval <- 5                             #Length of each age interval

#The transition rates are assumed to change at ages 5, 10, 15, 20, and 25.
#The variable TimeSpentInState contains the discounted time spent in each state.

for (i in 1:6){                           #For each interval with constant rates
#In an empirical application, the following 2 lines should be replaced by a calculation
of the
#age-specific transition rates.
Q[] <- runif(n^2)                         #Generate random matrix of transition
rates
diag(Q)<- diag(Q)-apply(Q,1,sum)          #Correct diagonal values
```

```
#Determine the length of time until the next change in transition rates, or the time
horizon.
if (Interval*i<T && i<6) CurrentInterval <- Interval
else CurrentInterval <- max(T-Interval*(i-1), 0)
#Update the discounted time spent in each state (see equation 7)
TimeSpentInState <- TimeSpentInState + exp(-r*(i-1)*Interval) * cbind(1,ZeroVector)
%*%
expm(CurrentInterval*rbind(cbind(0,InitStateDistr),cbind(t(ZeroVector), Q-r*I))) %*%
rbind(ZeroVector,I)
#Update the state distribution of the cohort
InitStateDistr <- as.matrix(InitStateDistr %*% expm(CurrentInterval*Q))
```

REFERENCES

1. Beck JR, Pauker SG. The Markov process in medical prognosis. Medical Decision Making. 1983;3(4):419-58.

2. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Medical Decision Making. 1993 Oct-Dec;13(4):322-38.

3. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. PharmacoEconomics. 1998 Apr;13(4):397-409.

4. Hunink MGM, Glasziou PP, Siegel JE, Weeks JC, Pliskin JS, Elstein AS, et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge, UK: Cambridge University Press 2001.

5. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. Medical Decision Making. 2011 Jan-Feb;31(1):10-8.

6. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Economics. 2002 Jan;11(1):33-42.

7. Naimark DM, Bott M, Krahn M. The half-cycle correction explained: two alternative pedagogical approaches. Medical Decision Making. 2008 Sep-Oct;28(5):706-12.

8. Barendregt JJ. The half-cycle correction: banish rather than explain it. Medical Decision Making. 2009 Jul-Aug;29(4):500-2.

9. Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. Journal of the Operational Research Society. 2007;58:168-76.

10. Cooper K, Brailsford SC, Davies R, Raftery J. A review of health care models for coronary heart disease interventions. Health Care Management Science. 2006 Nov;9(4):311-24.

11. Kim SY, Goldie SJ, Salomon JA. Exploring model uncertainty in economic evaluation of health interventions: the example of rotavirus vaccination in Vietnam. Medical Decision Making. 2010 Sep-Oct;30(5):E1-E28.

12. Varas-Lorenzo C, Maguire A, Castellsague J, Perez-Gutthann S. Quantitative assessment of the gastrointestinal and cardiovascular risk-benefit of celecoxib compared to individual NSAIDs at the population level. Pharmacoepidemiology and Drug Safety. 2007;16(4):366-76.

13. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. Health Economics. 2003 Oct;12(10):837-48.
14. Elbasha EE, Szucs T, Chaudhary MA, Kumar RN, Roediger A, Cook JR, et al. Cost-effectiveness of raltegravir in antiretroviral treatment-experienced HIV-1-infected patients in Switzerland. HIV Clinical Trials. 2009 Jul-Aug;10(4):233-53.
15. Hazen GB. Factored stochastic trees: a tool for solving complex temporal medical decision models. Medical Decision Making. 1993 Jul-Sep;13(3):227-36.
16. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs Biomedicine. 1985 May;20(1):79-93.
17. Welton NJ, Ades AE. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: uncertainty propagation, evidence synthesis, and model calibration. Medical Decision Making. 2005 Nov-Dec;25(6):633-45.
18. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. British Journal of Cancer. 1954 Mar;8(1):1-12.
19. Uhry Z, Hedelin G, Colonna M, Asselain B, Arveux P, Rogel A, et al. Multi-state Markov models in cancer screening evaluation: a brief review and case study. Statistical Methods in Medical Research. 2010 Oct;19(5):463-86.
20. Satten GA, Longini IM. Markov chains with measurement error: estimating the 'true' course of a marker of the progression of human immunodeficiency virus disease. Applied Statistics. 1996;45(3):275-309.
21. Miller DK, Homan SM. Determining transition probabilities: confusion and suggestions. Medical Decision Making. 1994 Jan-Mar;14(1):52-8.
22. Singer B, Spilerman S. The representation of social processes by Markov models. American Journal of Sociology. 1976;82(1):1-54.
23. Israel RB, Rosenthal JS, Wei JZ. Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. Mathematical Finance. 2001;11(2):245-65.
24. Cox DR, Miller HD. The Theory of Stochastic Processes. London: Chapman and Hall 1965.

25. Moler C, Van Loan C. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Review. 2003;45(1):3-49.

26. Van Loan C. Computing integrals involving the matrix exponential. IEEE Transactions on Automatic Control. 1978;23(3):395 - 404.

27. Tataru P, Hobolth A. Comparison of methods for calculating conditional expectations of sufficient statistics for continuous time Markov chains. BMC Bioinformatics. 2011;12:465.

28. Castella F, Dujardin G, Sericola B. Moments' analysis in homogeneous Markov reward models. Methodology and Computing in Applied Probability. 2009;11(4):583-601.

29. Telek M, Horváth A, Horváth G. Analysis of inhomogeneous Markov reward models. Linear Algebra and its Applications. 2004;386:383–405.

30. Hawkins N, Sculpher M, Epstein D. Cost-effectiveness analysis of treatments for chronic disease: using R to incorporate time dependency of treatment response. Medical Decision Making. 2005 Sep-Oct;25(5):511-9.

31. Asmussen S, Nerman O, Olsson M. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics. 1996;23(4): 419-41

32. Ross SM. Introduction to probability models. Amsterdam: Academic Press 2007.

33. Neuts MF. Matrix-geometric solutions in stochastic models, an algorithmic approach. Baltimore: John Hopkins University Press 1981.

34. Asmussen S. Applied probability and queues. Berlin: Springer 2003.

35. Fackrell M. Modelling healthcare systems with phase-type distributions. Health Care Management Science. 2009 Mar;12(1):11-26.

36. Thümmler A, Buchholz P, Telek M. A novel approach for phase-type fitting with the EM algorithm. IEEE Transactions on Dependable and Secure Computing. 2006;3(3):245-58.

37. Toy M, Redelop WK, Veldhuijzen IK, Richardus JH, Schalm SW. Modelling age-specific health gain and costs of antiviral therapy for active chronic hepatitis B. In submission.

38. Toy M, Veldhuijzen IK, de Man RA, Richardus JH, Schalm SW. Potential impact of long-term nucleoside therapy on the mortality and morbidity of active chronic hepatitis B. Hepatology. 2009;50(3):743-51.

39. Levy AR, Kowdley KV, Iloeje U, Tafesse E, Mukherjee J, Gish R, et al. The Impact of Chronic Hepatitis B on Quality of Life: A Multinational Study of Utilities from Infected and Uninfected Persons. Value in Health. 2008 2008/6//;11(3):527-38.

40. Dutch Health Care Insurance Board (CVZ).

41. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine. 2007;26(11):2389-430.

42. Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. Journal of the American Statistical Assocation. 1985;80(392):863-71.

43. Groot Koerkamp B, Weinstein MC, Stijnen T, Heijenbrok-Kal MH, Hunink MG. Uncertainty and patient heterogeneity in medical decision models. Medical Decision Making. 2010 Mar-Apr;30(2):194-205.

44. Stout NK, Knudsen AB, Kong CY, McMahon PM, Gazelle GS. Calibration methods used in cancer simulation models and suggested reporting guidelines. PharmacoEconomics. 2009;27(7):533-45.

45. Tan SYGL, Van Oortmarssen GJ, Piersma N. Estimating Parameters of a Microsimulation Model for Breast Cancer Screening Using the Score Function Method. Annals of Operations Research. 2003;119:43-61.

46. Rutter CM, Miglioretti DL, Savarino JE. Bayesian calibration of microsimulation models. Journal of the America Statistical Assocation. 2009 Dec 1;104(488):1338-50.

47. Kong CY, McMahon PM, Gazelle GS. Calibration of disease simulation model using an engineering approach. Value in Health. 2009 Jun;12(4):521-9.

Letter re: Cost-effectiveness of pertussis booster vaccination in the Netherlands

James F O'Mahony, MA.

Vaccine 30(50).

To the editor, - In their recent analysis of pertussis booster vaccination Rozenbaum et al. note the surprising finding that the vaccine is more cost-effective when costs and effects are discounted than when undiscounted (1). This result is indeed surprising. Preventative interventions typically impose costs in the present and yield health benefits and treatment cost savings in the future. Accounting for time preference by discounting reduces the value of these future benefits, while costs incurred in the present remain unchanged, typically resulting in less favourable cost-effectiveness ratios compared to undiscounted outcomes (2). Consider this the *conventional effect* of discounting.

A possible explanation for Rozenbaum et al.'s surprising finding lies with Dutch discount rates. Rozenbaum et al. apply differential discount rates of 4% & 1.5% for costs and health effects respectively, in accordance with Dutch cost-effectiveness analysis guidelines (3). The differential between the rates accounts for anticipated growth in the value of health overtime (4). Consequently, the later an intervention occurs relative to the discount year (the point in time to which costs and effects are discounted back to), the more cost-effective it will be, all else equal. This is because differential discounting deflates the cost-effectiveness ratio to account for the implied growth in the value of health between the discount year and the future period in which the intervention occurs. Consider this the *growing value of health effect*, which is particular to differential discounting.

In Rozenbaum et al.'s example, the booster strategies are modelled at different ages in a single birth cohort. Presumably this implies vaccination at a range of time lags from the discount year. It is possible that the growing value of health effect is dominating the conventional effect of discounting in this example, leading to the more favourable cost-effectiveness ratio under discounting. It would be interesting to know if Rozenbaum et al.'s strange result is still observed when costs and effects are discounted at an equal rate or if the discount year is changed to the year of booster vaccination.

A simple numerical example can illustrate how differential discounting may lead cost-effectiveness ratios more favourable than under zero discounting. Consider a hypothetical vaccination costing €60 per recipient. Assume an average QALY gain per recipient of 0.01 QALYs per year, sustained for three years from the date of vaccination. Assume the vaccine is equally costly and effective in children aged 2 and 12. The vaccine's undiscounted cost-effectiveness ratio (CER) is €2,000/QALY. Under differential discounting at the Dutch rates of 4% & 1.5% for costs and effects respectively, the CER of vaccination at age 2 is €2,030/QALY when costs and effects are discounted to the year of vaccination. Vaccination at age 12 also has a CER of €2,030/QALY when costs and effects are discounted to the year of vaccination. However, if the cost and effects of vaccination at age 12 are discounted back to age 2, then the CER of vaccination is €1,590/QALY. Consequently, the discounted CER is more favourable than the undiscounted CER.

This simple example illustrates the need for caution when interpreting the results of a single birth cohort model under differential discounting. Under equal discounting the cost-effectiveness ratio of the vaccine in a child currently aged 12 is the same as vaccinating a 2 year old in 10 years time, all else equal. However, under differential discounting the implied increase in the value of health means that vaccinating a 2 year old in 10 years time will be more cost-effective than a current 12 year old, assuming all else equal including the discount year. Adjusting the discount year to the year of the intervention is required if a single cohort model is to be representative of cost-effectiveness at different ages. Other examples of the practical problems of interpreting cost-effectiveness estimates under differential discounting have been noted elsewhere (5).

REFERENCES

1. Rozenbaum MH, De Cao E, Postma MJ. Cost-effectiveness of pertussis booster vaccination in the Netherlands. Vaccine 2012;DOI: 10.1016/j.vaccine.2012.06.026.

2. Lipscomb J, Weinstein MC, Torrence GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-effectiveness in Health and Medicine. Oxford: Oxford University Press; 1996. p. 214-46.

3. CVZ. Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie [in Dutch]. Diemen: College voor Zorgverzekeringen; 2006.

4. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Econ 2001;10(7):587-99.

5. O'Mahony JF, de Kok IMCM, van Rosmalen J, Habbema JDF, Brouwer W, van Ballegooijen M. Practical implications of differential discounting in cost-effectiveness analyses with varying numbers of cohorts. Value Health 2011;14(4):438-42.

Letter re: Sharp et al (2012) Cost-effectiveness of population-based screening for colorectal cancer

James F O'Mahony, MA.

British Journal of Cancer 108(5).

Sharp et al.'s cost-effectiveness analysis (CEA) of colorectal screening in Ireland sees the academic publication of a CEA published by Ireland's Health Information and Quality Authority (HIQA) in 2009 [1,2]. Both publications deserve comment regarding the choice of alternatives compared and the confusion of average and incremental cost-effectiveness ratios (ACERs and ICERs respectively).

First, regarding the number of alternatives compared. Both analyses describe a core of three screening strategies, biennial guaiac-based fecal occult blood testing between ages 55-74, biennial faecal immunochemical testing (FIT) between 55-74 and flexible sigmoidoscopy at 60, and five additional strategies with alternative age ranges. No screening frequency other than biennial is considered. No start age below 55 or stop age above 74 is considered. No FIT test quantitative cut-off level other than of 100 ng/ml haemoglobin is considered. The primary cost-effectiveness results reported in Sharp et al.'s abstract are for the three core strategies.

The total of eight screening strategies is a relatively narrow set of alternatives. The core of three is even narrower. No rationale is given for the consideration of three core strategies or their significance relative to the others considered. The choice of a narrow range of alternatives is potentially problematic, as it can result in the omission of other relevant alternatives that may be more cost-effective. Furthermore, the omission of relevant alternatives can bias the ICERs of the included strategies downwards.

The narrow set of alternatives also has implications for uncertainty analysis. Sharp et al. provide some summary estimates of cost-effectiveness acceptability curves. They report that FIT between 55-74 would have a 99% probability of being cost-effective at a threshold of €14,000/quality adjusted life year (QALY). However, this probability is not very meaningful, as no strategy more intensive than FIT between 55-74 is analysed: the apparent confidence in cost-effectiveness is due to the omission of relevant alternatives rather than a high degree of certainty.

Sharp et al. cite Wilschut et al.'s analysis of alternative FIT cut-off levels, which found all strategies with a cut-off of 50ng/ml to dominate all others [3]. They state that sensitivity is not a major driver of cost-effectiveness and that their conclusions would not change if they repeated their analysis in light of this recent evidence. This is doubtful. If on repeating the analysis Sharp et al. also found that a lower FIT cut-off dominates in all strategies, then they would not conclude that FIT screening with a 100 ng/ml cut-off is cost-effective. Wilschut et al.'s results suggest that Ireland's population colorectal screening programme will be unnecessarily inefficient unless the cut-off level is adjusted downwards from the 100 ng/ml assessed by Sharp et al.

Second, regarding the confusion between ACERs and ICERs. The cost-effectiveness ratios reported in Table 2 and the primary results reported in the abstract are described as ICERs, but should be described as ACERs, as they are the ratio of costs and effects of each alternative relative to no screening. The received interpretation of ICERs is given by Siegel et al., whereby ICERs should be estimated as the incremental cost-effectiveness only between those strategies not subject to dominance or extended dominance, i.e. those on the efficient frontier [4]. ICERs not ACERs are the appropriate metric for determining cost-effectiveness [5].

The final paragraph of the results section does correctly report the ICERs for the efficient alternatives, but these are not the ICERs reported in the abstract or Table 2. This distinction matters. The ACER of FIT between 55-74 erroneously reported in the abstract and Table 2 as an ICER is approximately €1,700/QALY, while the correct ICER relative to FIT between 55-64 is €3,200/QALY: the difference is almost a factor of two. Furthermore, reporting an ICER for gFOBT implies it is efficient, whereas it is not, as it is dominated by other strategies.

The consequence of this confusion in Sharp et al. and HIQA's CEA is an abundance of irrelevant cost-effectiveness ratios being reported as ICERs. Both publications only needed to report three ICERs; one each for the three efficient strategies. Reporting superfluous cost-effectiveness ratios as ICERs can only confuse decision makers,

especially when multiple ICERs are reported for the same strategy and ICERs are reported for inefficient strategies. This confusion of ICERs and ACERs has been further compounded by HIQA, as the 2009 CEA is used as an illustrative example of the reporting of cost-effectiveness estimates in their methods guidance [6].

The number of alternatives considered within CEAs will be bound by data and other constraints. However, unnecessary constraints on the number of alternatives considered should not be imposed, such as the choice of three core strategies in this case. Analysts should provide decision makers with the correct interpretation of their cost-effectiveness estimates and avoid unnecessary confusion.

REFERENCES

1. Sharp L, Tilson L, Whyte S, O'Ceilleachair A, Walsh C, Usher C, 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. British Journal of Cancer 2012;106(5):805-16.

2. HIQA. Health technology assessment of a population-based colorectal cancer screening programme in Ireland. Dublin: Health Information and Quality Authority; 2009.

3. Wilschut JA, Hol L, Dekker E, Jansen JB, van Leerdam ME, Lansdorp-Vogelaar I, et al. Cost-effectiveness Analysis of a Quantitative Immunochemical Test for Colorectal Cancer Screening. Gastroenterology 2011;141(5):1648-55.

4. Siegel JE, Weinstein MC, Torrance GW. Reporting Cost-Effectiveness Studies and Results. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996. p. 276-303.

5. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. JAMA Journal of the American Medical Association 1996;276(16):1339-41.

6. HIQA. Guidelines for the Economic Evaluation of Health Technologies in Ireland Dublin: Health Information and Quality Authority; 2010.

Using Growth in the Cost-Effectiveness Threshold to Inform the Differential between the Discount Rate on Costs And Health Effects: Reason To Be Cautious?

James O'Mahony [1,2]
omahonj1@tcd.ie

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands

ABSTRACT

BACKGROUND: Differential discounting of the costs and effects of healthcare interventions has been extensively debated. Prior to a recent publication by Claxton et al. some considered differential discounting to be justified by a growing value of health, whereby the discount rate for health effects equalled the discount rate on costs minus the approximate growth in the value of health. Estimates of the growth in the value of health rest on the income and health elasticities of utility and income and health growth rates, all of which are uncertain. Consequently, such estimates of the appropriate differential are speculative at best. The current discount rates for costs and effects in the Netherlands are 4 and 1.5 % respectively, implying a differential of 2.5 %. Claxton et al. have now shown that the discount differential used in cost-effectiveness analysis should approximate the annual rate of growth of the cost-effectiveness threshold. **ANALYSIS:** This link to threshold growth provides a more immediate alternative by which to empirically determine the appropriate differential. While cost-effectiveness thresholds are often not made explicit, there is no evidence that they have risen in recent years. Indeed, there are reasons to expect thresholds not to grow or possibly even decline, despite positive income growth. **CONCLUSION:** Existing evidence and expectations of threshold growth lead to conservative estimates of the discounting differential of at or near zero. The current Dutch differential of 2.5 % is probably too high. A reconsideration of the current Dutch discount rates is required to account for expectations of threshold growth.

1. Introduction

The debate over differential discounting gained considerable clarity with a recent publication by Claxton et al [1]. They show that the appropriate discount rates for the economic analysis of healthcare depend on whether cost-benefit or cost-effectiveness analysis (CEA) is used. They demonstrate that growth in the value of health should only determine the discount differential if cost-benefit analysis is used. However, CEA is more commonly used in the economic evaluation of healthcare. Claxton et al. show that when using CEA, the differential should approximate the annual growth in the cost-effectiveness threshold. This distinction matters, as our expectations of threshold growth are arguably clearer than growth in the value of health. Furthermore, growth in the threshold is, in principle, more directly observable than growth in the value of health.

Table 1. Discount rates in countries using differential discounting

	Costs, %	Effects, %	Differential, %
The Netherlands	4	1.5	2.5
Belgium	3	1.5	1.5
England & Wales*	6	1.5	4.5

* Prior to 2004

While differential discounting is still subject to extensive theoretical debate, it has been implemented in several countries (Table 1). The discount rates in table 1 were adopted under the assumption that the differential should approximate the growth in the value of health, not the growth in the threshold as is now understood to be appropriate in CEA

Box 1. Threshold Growth and the Discount Differential

The relationship between the growth in the cost-effectiveness threshold and the discounting differential can be shown simply. Consider an intervention implemented one year after the discount year (period 2) that is just cost-effective; that is, its ratio of cost to effects equals the threshold in that period, k_2. The annual growth in k is g_k, so k_2 can be written in terms of g_k and k_1 (1).

$$\frac{C}{E} = k_2 = k_1(1+g_k) \quad (1)$$

$$\frac{C/(1+d_c)}{E/(1+d_e)} = k_1 \quad (2)$$

$$d_e = \frac{1+d_c}{1+g_k} - 1 \approx d_c - g_k \quad (3)$$

C costs
E effects
d_c costs discount rate
d_e health effects discount rate
k cost-effectiveness threshold
g_k annual threshold growth

If the intervention is marginally cost-effective in period 2, it should also appear marginally cost-effective when assessed in advance in period 1. Therefore, when the costs and effects of implementation in period 2 are discounted to period 1, the intervention should also just equal the threshold in period 1, k_1 (2). From (1) and (2) we can solve for the discount rate applied to health effects in terms of the discount rate for costs and threshold growth rate. We find that the discount rate for effects is approximately the discount rate for costs minus the growth in the threshold (3).

2. Estimating Growth in the Value of Health

Income growth is anticipated to lead to a growing value of health. While the marginal utility from both growing income and health is anticipated to fall, the marginal utility from income is assumed to fall faster than that of health, resulting in an increasing value of health. However, estimates of the growth in the value of health require estimates of growth in income and health as well as the elasticity of utility with respect to both income and health. Each of these elements are highly uncertain and attempts to estimate growth in the value of health have been described as problematic [2]. Given such uncertainty, estimates of the growth in the value of health, such as that of Klok et al. for the Netherlands [3], can only be considered speculative.

3. Threshold Growth

If the threshold is assumed not to be chosen directly, but rather implied by the healthcare budget constraint and current healthcare technology [4], there are a number of reasons why it might not increase despite income growth. Increases in productivity, cost reductions and disinvestment of less efficient technologies are all anticipated to reduce the threshold, as they would improve the cost-effectiveness of the marginal intervention [5]. Similarly, an increase in healthcare demand would reduce the threshold. Consequently, as Paulden and Claxton note, while the threshold will expand with budget growth all else equal, there are a number of countervailing reasons for it not to grow or possibly fall [6].

While thresholds of €20,000 and €80,000 per QALY are commonly cited in Dutch CEAs, the Netherlands has no official threshold [7]. Consequently, it is impossible to meaningfully claim the Dutch threshold has grown or not. While this uncertainty also applies to many other countries with no official thresholds, the National Institute for Health and Clinical Excellence (NICE) in England and Wales uses a threshold range of £20-30,000 per QALY and has done so for almost a decade [5]. Although the use of a threshold range makes it harder to determine if the threshold has changed, NICE's threshold does not appear to have grown.

Both the reasons to expect low threshold grow and NICE's static threshold range support expectations of low or no threshold growth. Indeed, Paulden and Claxton have suggested that NICE and the Canadian Agency for Drugs and Technology in Health should adopt common discount rates, reflecting their expectation of no threshold growth [8]. While the lack of official thresholds in the Netherlands confounds attempts to estimate threshold growth, it is probable that any threshold would grow at less than the 2.5 % implied by the current discount rates.

4. Consequences of Incorrect Differentials

The consequences of adopting too high a discounting differential can be illustrated by example. Consider two types of interventions; those that yield health benefits shortly after their implementation (proximate interventions) and those with benefits after a considerable delay (distant interventions). Assume that the discount rate for costs is correct, but that the differential is too large: implying the discount rate on effects is too low. Too low a health discount rate causes relatively more distant interventions than proximate interventions to be considered cost-effective than would be the case had the correct health discount rate been applied. In the case of a fixed healthcare budget, this implies that the mix of interventions will not be optimal. In the case of an exogenously determined threshold and no budget constraint, both the intervention mix will be wrong and the amount of spending will be too high.

A recent CEA of Human Papillomavirus vaccination of 12 year old girls by de Kok et al. provides a good example of the potential influence of discounting on decision making [9]. They report the costs-effectiveness of vaccination in the Netherlands under both the Dutch discount rates of 4 and 1.5 % and a common rate of 3 % to be €19,700 and €53,500 per QALY respectively. Under the differential rates, this preventative intervention falls just under the commonly used lower threshold of €20,000. However, the relatively low common discount rate of 3 % results in a cost-effectiveness estimate well in excess of the threshold.

5. Recommendations

The current Dutch discount rates need to be reconsidered to ensure their consistency with the current understanding of the discounting differential within CEA. The lack an official cost-effectiveness threshold in the Netherlands is therefore problematic. In this respect, the recent recommendation by Boersma et al. that an official threshold be established is welcome [7]. Such a reconsideration of discounting should at least address how a Dutch cost-effectiveness threshold might grow in principle, even if no threshold will be established in practice.

More broadly, CEA advisory bodies should consider the theoretical arguments a empirical evidence regarding discounting very carefully. This is important due t sensitivity or cost-effectiveness estimates to discounting assumptions and the implications for the mix and volume of health spending.

References

1. Claxton K, Paulden M, Gravelle H, et al. Discounting and decision making in the economic evaluation of health care technologies. Health Economics 2010;doi: 10.1002/hec.1612.
2. Gravelle H, Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. Health Economics 2001;10:587-99.
3. Klok RM, Brouwer WBF, Annemans LJP, et al. Towards a healthier discount procedure. Expert Review of Pharmacoeconomics and Outcomes Research 2005;5:59-63.
4. Culyer A, McCabe C, Briggs A, et al. Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. 2007;12:56-58.
5. McCabe C, Claxton K, Culyer AJ. The NICE Cost-Effectiveness Threshold: What it is and What that Means. PharmacoEconomics 2008;26:733-44.
6. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. Centre for Health Economics Research Paper 53. York: University of York, 2009.
7. Boersma C, Broere A, Postma MJ. Quantification of the Potential Impact of Cost-effectiveness Thresholds on Dutch Drug Expenditures Using Retrospective Analysis. Value in Health 2010;13:853-56.
8. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. CADTH 2010 Symposium Halifax, 2010.
9. de Kok IMCM, van Ballegooijen M, Habbema JDF. Cost-effectiveness analysis of human papillomavirus vaccination in the Netherlands. Journal of the National Cancer Institute 2009;101:1083-92.

DOES DIFFERENTIAL DISCOUNTING ENHANCE DECISION MAKING?

James O'Mahony, MA [1,2]

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
omahonj1@tcd.ie

ABSTRACT

Objectives: To demonstrate that differential discounting can confuse comparisons of cost-effectiveness between interventions. In particular, to show that directly comparing cost-effectiveness ratios of different interventions from conventional cost-effectiveness models can result in misleading cost-effectiveness rankings under differential discounting.

Methods: A simple example of a comparison of the cost-effectiveness of two hypothetical interventions is used. The first intervention is a once-off vaccination, which imposes costs in one year alone per patient-cohort. The second intervention is a therapy for a chronic condition, which imposes costs for many consecutive years per patient-cohort. Both interventions are assessed on the conventional basis of a single cohort of patients, resulting in cost-effectiveness estimates for each intervention, which are then compared. An alternative comparison of cost-effectiveness is then made; it considers the opportunity cost of adopting the therapy for one cohort of patients in terms of the vaccination that could be provided for many cohorts over the same period. This alternative comparison yields a second pair of cost-effectiveness ratios which are compared to those from the conventional per cohort comparison.

Results: Comparing the interventions on the basis of the actual opportunity cost with the same distribution of spending over time shows the vaccination to be relatively more cost-effective than under the conventional per cohort analysis. Furthermore, the alternative comparison shows the two interventions to have the same relative cost-effectiveness as if the conventional per cohort comparison had been undertaken with discount rates unadjusted for growth in the cost-effectiveness threshold (equal discounting).

Conclusions: The debate over differential discounting has not considered the practical implications for decision making. Our analysis shows that accounting for threshold growth using differential discounting can make comparisons from conventional forms of cost-effectiveness analysis misleading as to which interventions are most cost-effective. Accounting for threshold growth using differential discounting is less relevant for decision making than previously understood.

1. INTRODUCTION

The models used in cost-effectiveness analysis typically simulate a single cohort of patients starting an intervention in the discount year [1]. Not all interventions are the same, some are concluded very quickly, others can take many years to complete. Some interventions deliver health effects immediately, while others such as prevention may only yield health effects long after implementation. This means that different interventions can have different distributions of costs and effects over time. Furthermore, differential discounting of costs and effects means that the timing of an intervention relative to the discount year will influence its cost-effectiveness [2]. Consequently, this calls attention to the need to understand the influence of model structure on results and what is really being compared between different analyses.

This poster shows how differential discounting can confuse comparisons of cost-effectiveness with an example of a comparison between a short intervention that is concluded within a year and a long intervention that takes 5 years to complete. We compare the interventions using both a conventional per cohort comparison and an alternative comparison based on the opportunity cost of one intervention in terms of the other. Our results show both the need to clarify which comparison is appropriate and that differential discounting does not alter the relative cost-effectiveness of different interventions in certain circumstances, and consequently may not enhance decision making.

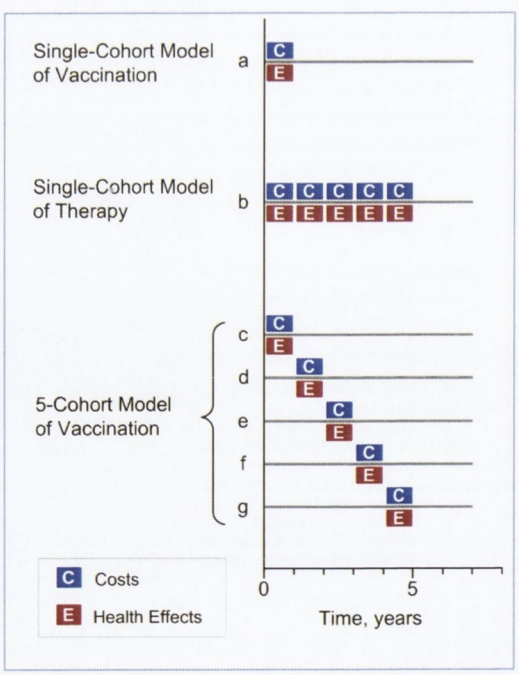
2. METHODS

We assume two interventions. The first is a vaccination costing €400 that yields a health benefit of 0.02 QALYs in the year in which it is administered. The vaccine provides no protection against infection in any subsequent year, such as seasonal influenza vaccines against changing disease strains. We also assume a therapeutic intervention that costs €400 per year and yields health benefits of 0.0195 QALYs in each year it is received. The therapy is assumed to be a treatment for a chronic condition and the complete treatment course is assumed to be 5 years per patient cohort.

We first compare the interventions' cost-effectiveness using a conventional per cohort analysis, which compares one cohort of patients receiving the vaccination in the discount year to a cohort of patients starting the 5 years of therapy in the discount year. We then make an alternative comparison that considers how much vaccination could be provided over the next 5 years with the resources consumed by the therapeutic intervention. This comparison shows the opportunity cost of implementing the therapy in terms of vaccination foregone. In this case, the opportunity cost of the therapy is vaccination of 5 consecutive cohorts in each of the 5 years from the discount year onwards. Figure 1 illustrates the comparisons. The conventional per cohort comparison is shown by comparison of cohorts a and b. The matched opportunity cost comparison is shown by comparing cohort b with cohorts c to g.

We assume that costs and effects are discounted differentially at rates of 4% and 1.5% respectively, in accordance with the current Dutch cost-effectiveness analysis guidelines [3]. We also discount costs and effects at an equal rate of 4%. We show the average cost-effectiveness of each intervention based on the present value of costs and effects in the discount year under each of the two discounting assumptions.

Figure 1: Single cohort comparison and 5 cohort comparison



3. RESULTS

Table 1 below shows the undiscounted costs and effects of each intervention in each year under each comparison, as well as the present value of costs and effects in the discount year and the associated cost-effectiveness ratios under both differential and equal discounting. Under differential discounting the cost-effectiveness ratio of vaccination is €20,000 per QALY for one cohort. The therapeutic intervention has a superior cost-effectiveness ratio of €19,546 per QALY for one patient cohort. However, when matched on spending over time, where the vaccination is implemented in 5 consecutive patient cohorts, vaccination now has a superior cost-effectiveness ratio of €19,075 per QALY.

Under equal discounting vaccination has the same cost-effectiveness under either the single or the multiple cohort comparisons of €20,000 per QALY. The therapy has a cost-effectiveness ratio of €20,513 per QALY.

The relative cost-effectiveness of vaccination compared to therapy (in the matched opportunity comparison) is shown as the ratio of the cost-effectiveness estimates. In both the differentially and equally discounted results the ratio is the same at approximately 1.03.

Table 1: Cost and effects of vaccination and therapy under different discounting assumptions and model structures

	Vaccination: Single Cohort		Vaccination: 5 Cohorts		Therapy: Single Cohort	
Year	Costs, €	Effects, QALYs	Costs, €	Effects, QALYs	Costs, €	Effects, QALYs
1	400	0.02	400	0.02	400	0.0195
2			400	0.02	400	0.0195
3			400	0.02	400	0.0195
4			400	0.02	400	0.0195
5			400	0.02	400	0.0195
Differential discounting, 4 & 1.5%						
Present Value	400	0.0200	1852	0.0971	1852	0.0947
ACER, €/QALY		20,000		19,075		19,564
Ratio of ACERs, therapy/ vaccination						1.03
Equal discounting, 4 & 4%						
Present Value	400	0.0200	1852	0.0923	1852	0.0903
ACER, €/QALY		20,000		20,000		20,513
Ratio of ACERs, therapy/ vaccination						1.03

Box 1: Constant Relative Cost-Effectiveness and Discounting Assumptions

The fact that the relative cost-effectiveness of the interventions is no different in the equal and differentially discounted cases when the cohorts are matched for spending is notable. This could be interpreted as meaning that differential discounting would not change the relative ranking between interventions if they were compared on an equal opportunity cost basis.

The validity of this interpretation depends on the assumptions regarding how differential discounting is applied. Recent work has shown that differential discounting is justified if the cost-effectiveness threshold is expected to grow over time [4]. In the example here it is assumed that the discount rate on health effects is adjusted downward by the approximate threshold growth rate ($d_h = r_e - g_t$, where d_h is the discount rate applied to health effects, r_e is the rate of time preference for costs, and g_t is the annual growth rate of the cost-effectiveness threshold). Under this assumption the equivalence of the relative cost-effectiveness observed in this example would not hold if the vaccination's health effects were experienced in years following vaccination.

There may however, be good reason to challenge the assumption that $d_h = r_e - g_t$: it may be the case that rather than the discount rate on health effects being lower, the discount rate on costs should increase by approximately the rate of threshold growth so that $d_c = r_e + g_t$, where d_c is the discount rate applied to costs. The equivalence of relative cost-effectiveness does hold when health effects occur later than costs under these alternative discounting assumptions.

4. DISCUSSION

In the case of the conventional per cohort comparison the therapy appears more cost-effective than vaccination. However, when we consider the opportunity cost of the therapy in terms vaccination, it then appears less cost-effective than vaccination. The current methodological literature gives no clear indication which model structures and comparisons are appropriate or which comparison should be used to identify the most cost-effective intervention. However, the comparison based on matched resource use presented here is consistent with the economic concept of opportunity cost, and so may be the more appropriate basis for comparison. That the relative cost-effectiveness of the interventions is equal under equal and differential discounting in the matched comparisons suggests that differential discounting would not enhance decision making in this case.

An important qualification of this conclusion is that the conventional per cohort comparison may be appropriate in certain circumstances. In particular, if the therapy is considered one single intervention in which the effects in each year are not independent of the intervention's implementation in previous years, then the per cohort comparison may be most relevant. However, if the effects of each year of the therapy are independent of each other, then it may not be considered one intervention, but 5 repeated instances of the same intervention. In this case the matched comparison would be appropriate.

Further research will be necessary to clarify which comparisons are appropriate under differential discounting. The example presented here shows it is currently unclear how results should be interpreted under differential discounting. Consequently, differential discounting may confuse rather than enhance decision making.

References

1. Standaert B, Demarteau N, Talbird S, et al. Modelling the effect of conjugate vaccines in pneumococcal disease: Cohort or population models? Vaccine 2010;28:G30-G38.
2. Hoyle M, Anderson R. Whose Costs and Benefits? Why Economic Evaluations Should Simulate Both Prevalent and All Future Incident Patient Cohorts. Medical Decision Making 2010.
3. College voor Zorgverzekeringen. Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie. 2006.
4. Claxton K, Paulden M, Gravelle H, et al. Discounting and decision making in the economic evaluation of health care technologies. Health Economics 2010;20.

THE DISCRETE MODELLING OF INTERVENTIONS WITH CONTINUOUSLY VARYING COSTS AND EFFECTS: IMPLICATIONS FOR ICERS, CEACS AND EVPI

James O'Mahony, MA [1,2] & Joost van Rosmalen, PhD [2]

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland.
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
omahonj1@tcd.ie

ABSTRACT

Objectives: To show how the discrete modelling of continuously varying cost-effect relationships in cost-effective analysis (CEA) influences incremental cost-effectiveness ratios (ICERs), cost-effectiveness acceptability curves (CEACs) and the expected value of perfect information (EVPI). In particular, to show that each of these metrics is contingent on the discrete comparisons chosen within a CEA.

Methods: The cost-effectiveness of a hypothetical intervention with a continuous dose-response relationship is simulated. The cost-effectiveness of a small number of possible dose levels is simulated first. The analysis is then repeated a number of times, progressively increasing the number of possible doses and resulting combinations of costs and effects. For each run of the analysis ICERs are calculated for each dose level, a probabilistic sensitivity analysis is simulated, and CEACs and the EVPI are plotted.

Results: As the number of potential cost and effect combinations increases, the ICERs for each dose level increase, the CEACs fall towards zero and the EVPI both rises and changes from having sharp inflection points to being a smooth, upwards-sloping curve.

Conclusions: Many interventions demonstrate dose-response relationships, most of which are in principle continuous, even if doses are typically varied discretely. The continuously increasing intensity of interventions means the number of possible alternatives is infinite. The general conclusion from the analysis is that each of the metrics presented here are contingent on the discrete comparisons chosen within the analysis. The significance of this finding for CEACs depends on their interpretation, which varies in the literature. The significance for EVPI is that while it has previously been recognised that excluding relevant comparators can reduce the EVPI, including all theoretically relevant alternatives may be impossible. Further work may be required to understand this constraint on measuring the upper bound of the value of further research.

1. INTRODUCTION

CEAs typically assess a small number of treatment alternatives. However, in many cases there is a large or even infinite number of possible alternatives. This poster shows that ICERs, CEACs and the EVPI are contingent on the number of alternatives compared. As it is unclear if there is any objective way of specifying the number of alternatives assessed or the intervals between them, both CEACs and EVPI appear somewhat arbitrary.

The possibility for a large number of treatment alternatives considered here is not that there are many distinct treatments available for a given condition, but that the intensity of interventions can be varied. Although we typically consider discrete increments in dose levels, clearly doses can, in theory, be varied continuously in many cases. If the intensity of a intervention can be varied continuously and there is a continuous dose-response relationship, this implies a continuous relationship between cost and effects.

2. METHODS

We use a very simple model of a hypothetical intervention to illustrate the consequences of increasing the number of treatment alternatives. The intervention is assumed to have a continuous dose-response relationship and exhibit diminishing marginal effectiveness. This underlying dose-response relationship gives rise to a continuous linear costs-effects relationship, shown as the blue curve in figure 1. We model discrete numbers of treatment alternatives (cost and effects combinations) along this continuous function, increasing from 4 to 13 to 49 alternatives.

We simulate a probabilistic sensitivity analysis (PSA) by assuming normally distributed variation in costs and effects around each simulated point on the assumed cost-effects curve. The level of uncertainty is assumed common across alternatives. The CEACs and EVPI for each of the three comparisons are derived from the simulated PSA.

Box 1. Background to CEACs and EVPI

CEACs were initially devised in the context of comparisons between two interventions, but have since been applied to comparisons between multiple alternatives [1]. They are derived from probabilistic sensitivity analysis of cost-effectiveness estimates using the net benefit framework. The probability represented by the CEACs corresponds to the proportion of times each treatment alternative has the highest net benefit over all the repeated iterations of the PSA for a given value of the cost-effectiveness threshold. These probabilities are plotted against a range of threshold values to give the CEACs.

CEACs are recommended by the National Institute for Health and Clinical Excellence (NICE) and other CEA authorities as part of CEA submissions as measures of uncertainty in CEA. However, despite their adoption by CEA authorities they have also received criticism as being potentially misleading [2].

The application of EVPI methods to uncertainty in cost-effectiveness analysis is a closely related topic to CEACs [3]. Economic theory suggests that decision makers should be risk neutral regarding cost-effectiveness, and so be indifferent to uncertainty. The application of EVPI to cost-effectiveness analysis has been proposed as a way of understanding the implications of uncertainty. In particular, it can be used to show the upper bound of the value of further research to reduce uncertainty.

Figure 1: PSA Scatter and Efficient Frontier for 4 Treatment Alternatives

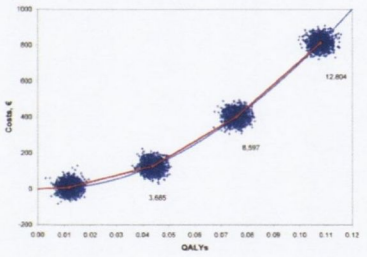


Figure 2: PSA Scatter and Efficient Frontier for 13 Treatment Alternatives

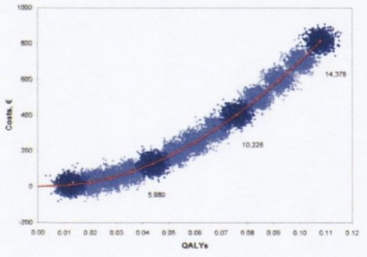


Figure 3: PSA Scatter and Efficient Frontier for 49 Treatment Alternatives

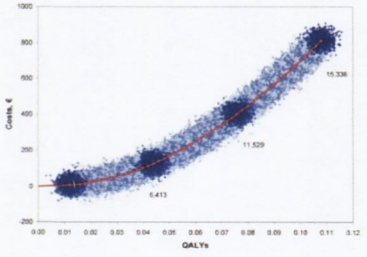


Figure 4: CEACs for 4 Treatment Alternatives

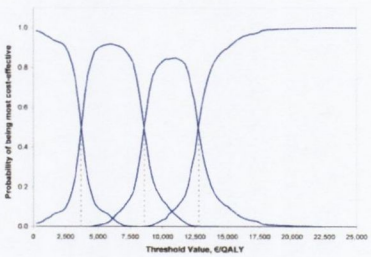


Figure 5: CEACs for 13 Treatment Alternatives

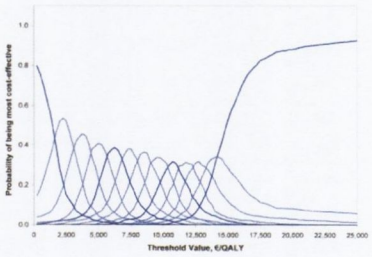
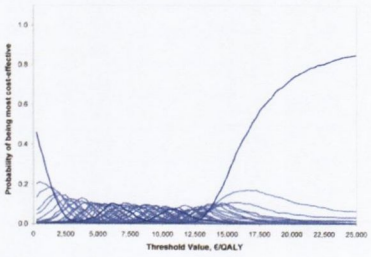
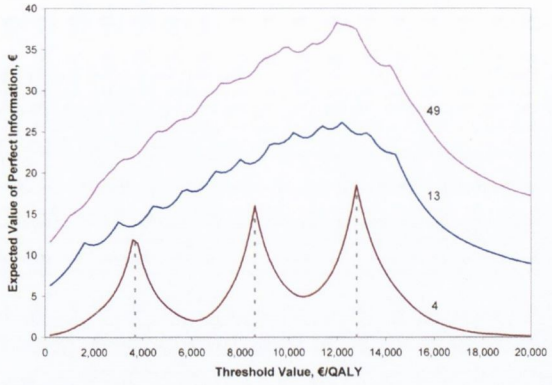


Figure 6: CEACs for 49 Treatment Alternatives



Please email for a colour PDF of the handout

Figure 7: EVPI curves for 4, 13 & 49 comparators



Box 2 Caveats and Extensions

The model used in this analysis is very simple and is not based on any empirical evidence of a real intervention. However, the model's simplicity is an advantage as it permits the clear illustration of a methodological problem with both CEACs and EVPI.

A major assumption of the model is that there is one continuous underlying cost-effectiveness frontier that is approximated with various degrees of accuracy with discrete points. This may not be the case in some diseases, as there may be several quite different treatments for a condition. However, continuous dose-response relationships may still exist within each of these treatment options, which would give rise to the same consequences illustrated here.

We assumed a degree of correlation of the costs and the effects between each alternative in our simulation of the PSA. Increasing this correlation reduces the rate at which the CEACs fall towards zero as the number of alternatives increases. Similarly, increased correlation moderates the increase in EVPI as the number of alternatives is increased. However, varying the degree of correlation does not change the overall general results presented here for CEACs and EVPI.

3. RESULTS

ICERs: Figures 1,2 & 3 show the PSA scatter and the cost-effectiveness frontier for the 4, 13 and 49 alternatives. The mean point of each cost-effect distribution is shown with the red diamonds, between each lies the cost-effectiveness frontier, shown with the red line. The PSA scatters for the original 4 dose alternatives in figure 1 are shown in dark blue in figures 2 and 3, while the additional alternatives are shown in lighter blue. The ICERs of the upper 3 of the 4 initial doses are shown in each figure. In each case the ICER increases as the number of alternatives increases. The ICER of the highest dose level increases from €12,804/QALY in the 4 dose comparison to €14,378 and €15,338/QALY in the 13 and 49 dose comparison respectively.

CEACs: Figures 4, 5 & 6 show the CEACs in each of the 3 simulations. In the 4 dose comparison, the 2 middle CEAC curves both attain probabilities exceeding 0.8. In the 13 dose comparison, the CEACs of the same middle 2 doses (marked in dark blue) only peak to probabilities of approximately 0.35 - 0.4. The same 2 CEACs in the 49 dose comparison only peak to probabilities of approximately 0.1. In general, the CEACs fall in probability as the number of comparators increases.

EVPI: The EVPI curves for the 3 simulations are shown together in figure 7. The EVPI curve for the 4 dose comparison is the lowest curve. It has 3 pronounced peaks, corresponding to the ICERs between the 4 doses, which also correspond to the level of the cost-effectiveness threshold where the CEACs cross in figure 4. The 13 dose EVPI curve is higher the 4 dose curve over all values of the threshold range. It still has some modest peaks, but the overall shape is that of a rising curve until a threshold value of approximately €12,000/QALY, at which point it slopes downwards. The 49 dose EVPI curve has a similar shape, although it lies above the 13 dose curve and is somewhat smoother.

4. DISCUSSION

The result that including more comparators results in higher ICERs is to be expected. As the number of comparators increases, the closest comparator becomes closer and the estimated ICER approaches the gradient of the underlying cost-effect relationship at that particular dose level. ICERs estimated from discrete comparisons are necessarily lower than those estimated as the gradient of the underlying cost-effectiveness function; by how much depends on the slope of the cost-effectiveness relationship and size of the interval between points in the discrete comparison.

The implications for CEACs of being partly determined by the number of comparators is that they then appear somewhat arbitrary. It is difficult to interpret the probability from a CEAC given the influence of the number of comparators: for example, it is uncertain if a low probability indicates that the estimates are highly uncertain, or if it is simply a consequence of a high number of comparators in the model.

The interpretation of the probability represented by CEACs varies in the literature. One interpretation is that it is the probability that the given intervention has the highest net benefit of all interventions compared at a given threshold [1]. This interpretation is consistent with our results. However, a broader interpretation of the probability is that it is simply the probability the given intervention is cost-effective at a given threshold [4]. That an intervention's probability of cost-effectiveness might vary with the number of comparators included in the analysis is not intuitive: our results are not consistent with this broader interpretation of CEACs. Consequently, one implication of our results is that clarification of the interpretation of CEACs is necessary.

The EVPI results are straightforward to interpret. When interventions are relatively close in the cost-effectiveness plane, the probability of making an error in the optimal choice is large, due to overlap of the net benefit distributions. Consequently, EVPI rises as the comparators become closer as more are included in the analysis. However, this rise in EVPI is anticipated to be to a limit, because the costs of making an error fall as the difference in net benefit between the optimal and non-optimal choice will diminish as the comparators become closer. Further work will be required to fully understand this limit and how it is approached with finite numbers of comparators. Like CEACs, the variation of EVPI with the number of comparators confuses its interpretation and makes it seem somewhat arbitrary.

5. CONCLUSION

Both CEACs and EVPI appear somewhat arbitrary given their variation with the number of alternatives compared within CEA. The difficulty in interpreting these metrics compromises their usefulness as objective measures of uncertainty in CEA. Further work will be required to clarify their interpretation.

References

1. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. Health Economics 2001;10.
2. Groot Koerkamp B, Hunink MGM, Stijnen T, et al. Limitations of Acceptability Curves for Presenting Uncertainty in Cost-Effectiveness Analysis. Medical Decision Making 2007;27:101-11.
3. Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation. Oxford: Oxford University Press, 2006.
4. Barton GR, Briggs AH, Fenwick EAL. Optimal Cost-Effectiveness Decisions: The Role of the Cost-Effectiveness Acceptability Curve (CEAC), the Cost-Effectiveness Acceptability Frontier (CEAF), and the Expected Value of Perfection Information (EVPI). Value in Health 2008;11:886-97.

How cost-effectiveness acceptability curves vary with the number of treatment strategies compared and why this compromises their usefulness

James O'Mahony, MA [1,2] Joost van Rosmalen, PhD [2] Myriam Hunink, MD, PhD [3,4]

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
[3] Department of Epidemiology, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
[4] Department of Health Policy and Management, Harvard School of Public Health, Boston, USA
omahonj1@tcd.ie

Purpose: To show that cost-effectiveness acceptability curves (CEACs) are contingent on the number of alternative treatment strategies compared and explain how this compromises their use as an objective measure of uncertainty in cost-effectiveness analysis (CEA). **Methods:** CEAs typically compare a finite number of treatment alternatives. However, in many cases the actual number of treatment alternatives is very large or infinite. In such cases the cost-effectiveness frontier is not composed of a discrete number of points in the cost-effectiveness plane, but can be continuous. We use the example of a hypothetical intervention with a continuous dose-response relationship to show how increasing the number of treatment alternatives compared influences the shape of CEACs. How these curves change depends on the correlation of uncertainty between treatment alternatives in the probabilistic sensitivity analysis used to derive the CEACs. We compare the cases of perfect and zero correlation and an intermediate case, in which the correlation between alternatives increases with their proximity in the cost-effectiveness plane. **Results:** In the case of zero correlation, increasing the number of treatment alternatives causes the CEACs to fall towards a probability of zero (Figures 2A & B). With perfect correlation, the curves lie at probability of zero and jump to a probability of one over the range of the cost-effectiveness threshold where the given intervention has the highest net benefit. As the number of alternatives included grows large, the portion of the CEAC lying at probability of one converges to a single spike. In the intermediate case, the CEACs may initially lie at probability of zero or one, as in the case of perfect correlation, but eventually fall towards zero as the number of alternatives grows large (Figures 2C & D). **Conclusions:** This analysis shows that CEACs are contingent on the number of treatment alternatives compared. Without an objective basis to choose the number alternatives or the increments between them, the resulting CEACs seem arbitrary in part. Consequently, the usefulness of CEACs as an objective measure of uncertainty is questionable when many treatment alternatives are possible.

Introduction

Cost-effectiveness acceptability curves (CEACs) are used to represent uncertainty in cost-effectiveness analysis (CEA).[1] CEACs are typically derived from probabilistic sensitivity analysis (PSA); they show the proportion of PSA iterations for which a given intervention has the maximum net (health or monetary) benefit over all alternatives considered for a range of cost-effectiveness threshold values.[2]

CEAs generally compare a finite number of intervention alternatives. These alternatives may be distinct interventions or simply varying amounts of the same intervention. For example, in the case of cervical cancer prevention we may compare Human Papillomavirus vaccination to cervical screening, or we may compare different frequencies of screening. While there may be a finite number of unique treatments for a given condition, the dosage of many interventions can vary continuously, meaning the potential number of treatment alternatives can be infinite.

This poster shows how CEACs can change with the number of alternatives considered. It draws attention to the influence of correlations between alternatives in the PSA on the change in CEACs with increased numbers of alternatives.

Figure 1A: 4 Treatment Alternatives & their PSA Scatters in the CE Plane

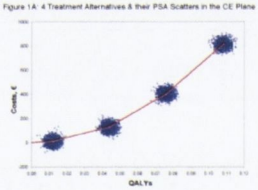
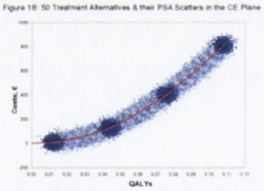


Figure 1B: 50 Treatment Alternatives & their PSA Scatters in the CE Plane



Methods

We consider a hypothetical intervention that can be delivered at different doses, resulting in different combinations of costs and health effects. We consider 4 and then 50 different doses. For simplicity, we choose mean cost-effect combinations for each dose level that describe an efficient frontier.

We simulate a PSA by assuming distributions of costs and effects around each mean value. We assume bivariate normal distributions of costs and effects of equal variance for all dose levels. No correlation is assumed between the uncertainty of costs and effects. However, we do make assumptions regarding the correlation of uncertainty of costs and effects between alternatives: first that there is no correlation and then that there is a high correlation between alternatives, with that correlation diminishing the further in the CE plane the alternatives are spread. CEACs are derived from the PSA under the 4 and then 50 alternatives and under the different correlation assumptions.

Results

The 4 and then 50 treatment alternatives and their simulated PSA scatters are shown in the cost-effectiveness plane in Figures 1A & B above. The red line marks the efficient frontier described by the mean points of costs and effects. Figure 1B shows the PSA scatters for the four original alternatives in dark blue, with those of the other 46 alternatives shown in lighter blues.

The consequences for CEACs of adding more alternatives under the assumption of uncorrelated alternatives are shown in Figures 2A & B, and under the assumption of high correlation in Figures 2C & D. In both cases adding more comparators results in CEACs falling towards zero, with the effect being much more pronounced when costs and effects are uncorrelated between alternatives.

Analysis

The inclusion of more alternatives has a pronounced effect on the CEACs when there is no correlation between alternative interventions, as the relative uncertainty between alternatives is high: adding more comparators greatly reduces the probability that any one given intervention will be most cost-effective at a given threshold value. Conversely, when costs and effects are highly correlated between alternative interventions, the relative uncertainty between alternatives is low. In this case, increasing the number of alternative interventions only modestly reduces the probability of an given intervention being most cost-effective at a given threshold.

Figure 2A: 4 Uncorrelated Alternatives

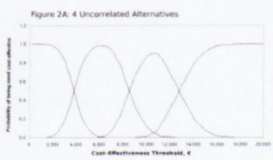


Figure 2B: 50 Uncorrelated Alternatives

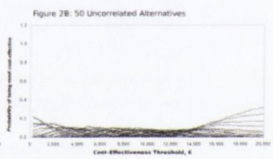


Figure 2C: 4 Correlated Alternatives

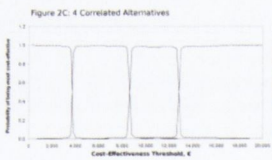
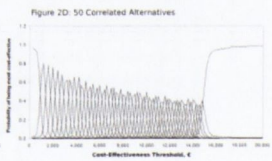


Figure 2D: 50 Correlated Alternatives



Discussion

The case of multiple uncorrelated treatment alternatives is unlikely to occur, as there are unlikely to be many unique treatments for a single condition with uncorrelated uncertainties of outcomes. Multiple possible alternatives are more likely to arise due to varying doses of one intervention, where costs and effects are likely to be highly correlated across dose levels. It is possible that this correlation will diminish as dosage levels diverge, as assumed in our example. So while a very significant flattening of CEACs with the inclusion of more alternatives is unlikely, it is still probable that CEACs will change with increased numbers of alternatives in many cases.

The interpretation of CEACs in the context of multiple alternatives is not intuitive. In particular, that the probability of any one strategy being cost-effective depends on the number of strategies compared makes CEACs seem arbitrary in part.

CEACs are highly dependent on the correlation of costs and effects between alternatives. If there is a lack of reliable data on these correlations, then the resulting CEACs may themselves be unreliable representations of uncertainty in CEA.

A limitation of our analysis is that we use a hypothetical intervention to illustrate our methodological point. However, such an example permits a very clear illustration of the influence of multiple possible alternatives on CEACs.

Conclusion

The contingency of CEACs on the number of alternatives considered when the set of possible alternatives is large makes them appear somewhat arbitrary. Consequently, their usefulness as an objective and intuitive measure of uncertainty in CEA is questionable.

References

1. van Hout et al (1994) Costs, Effects and C/E-Ratios Alongside a Clinical Trial: *Health Economics*
2. Fenwick et al (2001) Representing Uncertainty: The Role of Cost-Effectiveness Acceptability Curves. *Health Economics*

SHOULD WE AGGREGATE COST-EFFECTIVENESS OVER AN INTERVENTION'S ENTIRE IMPLEMENTATION LIFETIME?

James O'Mahony, MA [1,2]

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
omahonj1@tcd.ie

OBJECTIVES: Recent work has suggested that interventions' cost-effectiveness should be assessed over their entire lifetime of implementation, not just over the period of use for a single cohort as typically modelled (Hoyle and Anderson, Medical Decision Making, 2010; Hoyle, PharmacoEconomics, 2011). Such lifetime modelling can capture changes in costs and effects over time. These changes in costs and effects can result from price changes, disease dynamics or the application of differential discounting of costs and health effects. **METHODS:** Suggesting cost-effectiveness be assessed over an intervention's complete lifetime carries assumptions regarding the nature of the decision problem in healthcare resource allocation. In particular, it suggests resources be allocated on the basis of the total cost-effectiveness over all periods in which it is implemented. This lifetime perspective can conflict with the alternative perspective that resources be allocated on the basis of relative cost-effectiveness within each given period. We discuss a number of simple theoretical examples in which the rank ordering of cost-effectiveness of two interventions is different under the two perspectives. The examples include when the prices of interventions trend and have different expected lifetimes, when differential discounting is applied in certain circumstances, or simply when the price of only one intervention falls following patent expiry. **RESULTS:** These examples prompt us to consider which perspective is more appropriate. We argue that as health care resource allocation is an ongoing, repeated resource allocation problem, not one over a finite horizon, that the lifetime perspective is not appropriate. **CONCLUSION:** Advances in decision analytic modelling need to carefully reflect the actual nature of policy choices. The per-period perspective appears more appropriate to healthcare resource allocation problems than the total implementation lifetime perspective. However, the actual resource allocation process is likely to more complex than either perspective alone might suggest.

Introduction

Two recent and related papers by Hoyle and Hoyle and Anderson have argued that cost-effectiveness should be estimated over an intervention's entire expected implementation lifetime over all recipient cohorts.[1,2] The rationale for such implementation lifetime modelling given is that cost-effectiveness can change over the period of implementation and this should be accounted for in the present value of costs and effects when reimbursement decisions are taken. Both papers use the example of price reductions of pharmaceuticals following patent expiry. The implication being that there are interventions which are not cost-effective in current recipient cohorts, but will be cost-effective when sufficient future cohorts are included in the analysis, and so should be reimbursed now.

The objection to the proposed lifetime analysis presented here is, in simple terms, that if an intervention is not currently cost-effective, but will become so in the future, then reimbursement should be postponed until the intervention becomes cost-effective, rather than extending the analysis to include future cohorts and granting reimbursement now.

Factors Changing Cost-Effectiveness

Clearly a future reduction in the price of a drug following patent expiry will make it more cost-effective when that price reduction occurs. Shin et al. note the converse situation, when a reduction in the price of an existing comparator causes a new intervention to become less cost-effective.[3]

Hoyle and Anderson also note that cost-effectiveness will change over time when differential discounting is used. Differential discounting applies a lower discount rate to health effects than to costs to account for a growing value of health over time and is required in Belgium, the Netherlands and Poland. Differential discounting causes the cost-effectiveness of an intervention to improve in successive cohorts over time.[1,4]

Intervention effectiveness may also change over time, leading to changing cost-effectiveness. For example, in the case of infectious diseases, it may be some time before a vaccination programme yields herd immunity effects, but its cost-effectiveness will improve once these emerge. Conversely, the introduction of one intervention can have adverse effects on the cost-effectiveness of other complementary interventions. For example, the anticipated reduction in the incidence of cervical cancer following the adoption of vaccination against the Human Papillomavirus will cause the cost-effectiveness of cervical screening to deteriorate.

Examples of Changing Cost-Effectiveness Rank Order

The marked price reductions following patent expiry provide the most obvious example in which comparing interventions over their expected implementation lifetimes may give different cost-effectiveness estimates compared to implementation in the current cohorts alone. One intervention may be less cost-effective than another when compared in current cohorts of patients, but it may be more cost-effective when the analysis is extended to include future cohorts receiving a lower post-patent price.

Post-patent price reductions are examples that lead to abrupt changes in cost-effectiveness. Differential discounting causes gradual changes in cost-effectiveness over time. Adopting the lifetime perspective can cause the relative ranking of cost-effectiveness of two interventions to change from that observed when compared on the basis of current cohorts when differential discounting is applied. Despite the fact that differential discounting results in a common trend in cost-effectiveness in both interventions, differences in the implementation lifetimes between the two interventions can lead to a switch in the cost-effectiveness rank order.

Perverse Allocation of Consumer Surplus

The case of patent expiry provides a simple example of the counter-intuitive consequences of the lifetime perspective. Awarding reimbursement to an intervention which is not cost-effective at its current patent-protected price in anticipation of future post-patent price reductions rewards the patent holder an economic rent due to the consumer surplus generated by the future lower price, even though the current patent protected price generates no consumer surplus. Indeed, the patent holder is able to capture the entire economic rent, even though it may not produce the drug itself once the patent has expired. Furthermore, it is uncertain if the payer will reap that consumer surplus, as both the future price and use of the drug is uncertain. Consequently, reimbursement based on implementation lifetimes that account for post-patent prices would appear to be very generous to patent holders and leave payers bearing considerable risk.

Nature of the Decision Problem

Given the possible reversals in cost-effectiveness rank order of interventions between the conventional current cohort and the alternative implementation lifetime perspectives and possible perverse reimbursement decisions, it is important to consider which is more appropriate to healthcare decision making.

In most cases there is no need to accept or reject the use of an intervention over its entire implementation at one point in time. Rather, if an intervention is found not to be currently cost-effective and is not reimbursed, it is possible to review this decision in the future when prices or thresholds have changed.

Healthcare resource allocation is an ongoing, repeated resource allocation problem that is undertaken year after year: it is not once-off choice over a finite horizon. Consequently, it seems that assessing cost-effectiveness over short horizons with the possibility of review as necessary seems more appropriate than assessing interventions over their entire anticipated implementation lifetimes.

Caveats of Disinvestment and Cohort Separability

The decision problem may be more complicated when a review would suggest disinvestment rather than reimbursement. It may be difficult to withdraw an intervention once it is adopted, even if it is no longer cost-effective. Consequently, if an intervention is currently cost-effective, but is anticipated to become not cost-effective in the near future, then it may be necessary to withhold reimbursement or impose conditions on its use.

Some interventions may not be easily separable between cohorts, so postponing implementation to future periods may not be possible. For example, in the case of infectious disease control with herd immunity benefits, it might be necessary to implement it in current cohorts now when it is not cost-effective in order to build up herd immunity and achieve cost-effective implementation in the future. Similarly, in the case of capital interventions shared by generations of cohorts, it is not possible to selectively provide the intervention for some cohorts but not for others. In such cases where the separation of recipient cohorts is not possible, the entire implementation lifetime perspective may be more appropriate.

Conclusion

Anticipating future changes in cost-effectiveness can be useful. However, rather than assessing interventions over their entire implementation lifetime, it would be better to use predictions of changing future cost-effectiveness to identify when cost-effectiveness estimates should be reviewed. The lifetime perspective does not appear appropriate to most healthcare resource allocation problems. However, a implementation lifetime perspective may be appropriate in some particular cases.

References

1. Hoyle M, and Anderson R. Whose Costs and Benefits? Why Economic Evaluations Should Simulate Both Prevalent and All Future Incident Patient Cohorts. Med Decis Making 2010; 30(4).
2. Hoyle M. Accounting for the Drug Life Cycle and Future Drug Prices in Cost-Effectiveness Analysis Pharmacoeconomics 2011; 29 (1).
3. Shih YT, Han S, and Cantor SB. Impact of Generic Drug Entry on Cost-Effectiveness Analysis. Med Decis Making 2005 25(1).
4. O'Mahony JF, de Kok IMCM, van Rosmalen J, et al. Practical Implications of Differential Discounting in Cost-Effectiveness Analyses with Varying Numbers of Cohorts. Value in Health 2011: 14(4).

Revisiting HPV Vaccination:
Why Existing CEAs Underestimate the Vaccine's Cost-Effectiveness and Incorrectly Estimate its Threshold Price

James O'Mahony, MA [1,2]

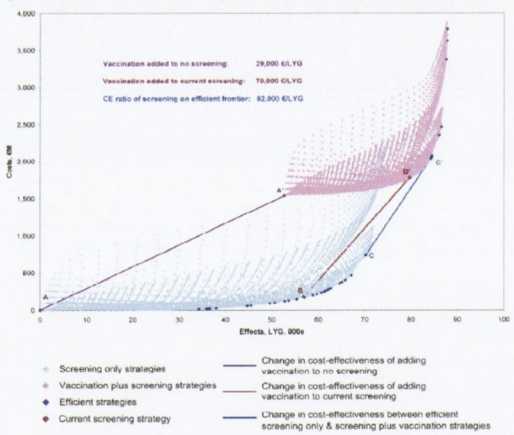
[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
omahonj1@tcd.ie

OBJECTIVES: Existing cost-effectiveness analyses (CEAs) of Human Papillomavirus (HPV) vaccination assume cervical screening remains unchanged. However, current screening intensities are unlikely to be cost-effective due to the likely reduction in disease incidence in vaccinated women. Therefore, reductions in screening intensity are probable. The cost-effectiveness attributable to vaccination varies with screening intensity. The assumption of unaltered screening leads to an underestimation of vaccine cost-effectiveness relative to when screening intensity is reduced. Furthermore, failure to consider other screening intensities yields an incomplete efficient frontier in the cost-effectiveness plane. This can lead to an incorrect estimate of the price at which vaccination becomes marginally cost-effective for a given cost-effectiveness threshold. **METHODS:** We review cost-effectiveness estimates for a wide range of screening only and vaccination plus screening strategies from a model used to estimate vaccine cost-effectiveness in the Netherlands. We indicate what comparison was used to estimate vaccine cost-effectiveness in previous studies, show what comparisons would be more appropriate and explain how these differ. **RESULTS:** We then show why the cost-effectiveness of adding vaccination to a given screening strategy is not the appropriate basis to determine if the vaccine is cost-effective or the threshold price. Rather, both should be determined by the ICER between the most costly efficient screening only strategy and the least costly vaccination plus screening strategy, even where this least costly vaccination plus screening strategy is not the optimal strategy for a given threshold. **CONCLUSIONS:** CEAs of HPV vaccination may no longer be policy or research priorities following widespread reimbursement and precipitous price reductions. However, the methodological issues raised here are pertinent to both any future CEA of an enhanced vaccine with protection against more HPV types and more generally to cases in which the cost-effectiveness of complementary interventions are not independent.

Introduction

Existing cost-effectiveness analyses (CEAs) of Human Papillomavirus (HPV) vaccination for the prevention of cervical cancer have not considered the consequences for optimal screening of vaccinated women as part of their analyses.[1] While it is recognised that screening may need to change for vaccinated women due to anticipated reductions in disease incidence,[2] this has not been considered in CEAs of vaccination. This has been considered a policy question that will not need to be answered until the current cohorts of vaccinated girls reach screening ages, which is at least a decade from now. This analysis shows how the cost-effectiveness of screening vaccinated women needs to be considered as part of the CEA of vaccination itself and therefore should be considered now.

Figure 1: Cost-effectiveness of screening only and vaccination plus screening strategies



Methods

We used the MISCAN microsimulation model of cervical cancer prevention developed at Erasmus MC, Rotterdam to simulate a range of screening only and vaccination plus screening strategies. The model used here is based on de Kok et al.'s CEA of HPV vaccination in the Netherlands and largely follows the assumptions described there.[3] Key parameters include a vaccine price of €118 per dose, 85% vaccine coverage and a discount rate of 3% for costs and health effects.

This analysis simulates not only the current screening strategy of screening every 5 years from ages 30 to 60, but also considers 10,800 alternative screening strategies, with different screening frequencies and screening start and stop ages.

The estimated costs and health effects, reported in terms of life years gained (LYG), of all strategies were plotted in the cost-effectiveness plane and the efficient frontier identified. We estimated the incremental cost-effectiveness of adding vaccination to the current screening strategy, to the no screening option and the incremental cost-effectiveness between the most intensive efficient screening only strategy and the least intensive vaccination plus screening strategy.

Results

The results of the CEA are plotted in the cost-effectiveness plane in Figure 1. The light blue points mark the unvaccinated strategies, while the pink mark the vaccination plus screening strategies. The efficient strategies are marked with the dark blue diamonds.

The difference in cost-effectiveness of adding vaccination to a no screening strategy is marked with the purple line and has a CE ratio of approximately €29,000 /LYG. The difference in cost-effectiveness of adding vaccination to the current screening strategy is marked with the dark red line and has CE ratio of approximately €70,000 /LYG. The change in costs-effectiveness between the last efficient screening only strategy and first vaccination plus screening strategy is shown with the blue line and has a CE ratio of approximately €92,000 /LYG.

Analysis

De Kok et al. and others in the literature have assumed current screening intensity will be maintained when estimating the cost-effectiveness of HPV vaccination. This corresponds to the cost-effectiveness (CE) ratio represented by the red line between B and B' of €70,000 /LYG. However, this comparison is not the correct basis to determine the cost-effectiveness of vaccination.

The blue line of C to C' shows the cost-effectiveness of moving between the most intense efficient screening only strategy and the least intense vaccination plus screening strategy. For vaccination to be efficient and cost-effective, the cost-effectiveness threshold must be at least €92,000 /LYG, which is considerably higher than the CE ratio of adding vaccination to the current screening intensity. The analysis shows that it would be more efficient to first increase screening intensity beyond its current level at point B, rather than adopting screening. Only if the CE threshold is greater than €92,000 /LYG would it be cost-effective to adopt vaccination and move from C to C'. Note that the two efficient strategies at C and C' are not of the same screening intensity.

The threshold price of the vaccination can be derived from our analysis. Assuming for convenience that the CE threshold is €70,000 /LYG, then the threshold price implied by the conventional analysis would be the analysed price of €118 per dose. However, if we more correctly derive the threshold price from the CE ratio of moving from a screening only to a screening plus vaccination strategy, we find a lower price of approximately €90 a dose, or 75% of the conventionally derived threshold price.

Discussion

A complete appraisal of the cost-effectiveness of HPV vaccination requires consideration of the full range of cervical cancer prevention strategies in the CE plane, not just a comparison between the cost and effects of the current screening intensity with and without vaccination. A complete analysis is required because adding vaccination to the current strategy may not be efficient, as is the case in this example. A lower price is required for a switch between any screening only to a screening plus vaccination price to become cost-effective. An even lower price again would be required for adding vaccination to the current screening strategy to be a cost-effective strategy.

This case presents an interesting example of the difficulties of estimating the cost-effectiveness of two complementary interventions. In the case of screening, its intensity can be varied, changing its cost-effectiveness. In the case of vaccination, its price is a key variable determining cost-effectiveness.

Conclusions

Existing CEAs of HPV vaccination have not taken sufficient account of the potential for varying screening intensity to determine the correct cost-effectiveness estimate or threshold price of the vaccination. The intensity of screening of vaccinated women is not a policy question for the future, but is relevant now.

References

1. Koleva D, De Compadri P, Padula A, and Garattini L. Economic evaluation of human papilloma virus vaccination in the European Union; a critical review. Internal and Emergency Medicine 2011; 6(2).
2. Beutels P, and Jit M. A brief history of economic evaluation for human papillomavirus vaccination policy. Sexual Health 2010; 7(3).
3. de Kok IMCM, van Ballegooijen M, and Habbema JDF. Cost-Effectiveness of Human Papillomavirus Vaccination in the Netherlands. JNCI 2009; 101(15).

DIFFERENTIAL DISCOUNTING: QUESTIONING THE ASSUMPTION OF HEALTHCARE RESOURCE FUNGIBILITY OVER TIME

James O'Mahony, MA [1,2]

[1] Centre of Health Policy and Management, School of Medicine, Trinity College Dublin, Dublin, Ireland
[2] Department of Public Health, Erasmus Medical Centre, Erasmus University, Rotterdam, The Netherlands
omahonj1@tcd.ie

OBJECTIVES: Recent work on differential discounting of cost and health effects has reached a degree of consensus in a previously strongly divided debate. Put simply, it holds that the discount rate applied to health effects should equal the discount rate for costs, less the growth rate of either the cost-effectiveness threshold or the consumption value of health, depending on the objectives of the health system. Assuming positive growth in the threshold or the value of health, this implies the cost-effectiveness of preventative interventions improves relative to the situation under equal discounting. **METHODS:** We show how recent analyses of differential discounting implicitly assume healthcare funds to be completely fungible over time. This assumption is difficult to justify in the context of publically funded healthcare systems that exhaust budgets annually. Assuming funds are not fungible results in alternative differential discount rates: in this case, the discount rate on costs should be adjusted upwards by either the growth rate of the threshold or the consumption value of health. **RESULTS:** Under these discount rates, interventions that impose costs in future periods become more cost-effective relative to the situation under equal discounting, rather than those which yield health gains in the future. Indeed, the cost-effectiveness of preventative interventions that reduce future healthcare costs will deteriorate under such alternative differential discounting. Consequently, interventions' cost-effectiveness may differ greatly between the two differential discounting schemes. **CONCLUSION:** Cost-effectiveness estimates can be highly sensitive to discounting; therefore the theory underpinning discount rates needs to be robust. This analysis shows that the current understanding of differential discounting needs to be re-examined. CEA authorities in countries currently employing differential discounting such as Belgium and The Netherlands and those contemplating it such as England and Wales should consider these issues carefully.

What Claxton et al. Showed

Claxton et al. recently demonstrated the rationale for the differential discounting of costs and health effects under three alternative health economic decision rules.[1] Arguably, the most relevant of these decision rules for CEA is where the incremental cost-effectiveness ratio is compared to a cost-effectiveness threshold. Where the threshold, rather than representing any value of health, is the cost-effectiveness of the marginal intervention that can be afforded within the current healthcare budget. In this case, the differential between the discount rates for costs and effects approximately equals the threshold's growth rate:

$$d_h = r_h$$
$$d_c \approx r_h + g_k$$

Where d_h is the discount rate applied to health effects; r_h is the rate of time preference for health; d_c is the discount rate applied to costs; and, g_k is the threshold growth rate.

What was not Resolved

Claxton et al. did not resolve what r_h is, which is necessary in determining what discount rates to apply in CEA. However, r_h has been addressed in a subsequent paper by Paulden and Claxton.[2] They use a two-period model in which health spending is allocated between periods to maximise the present volume of health effects. The time preference for health, r_h, is revealed by the budget allocation between the two periods: which Paulden and Claxton show to be a function of market interest rates and growth in the cost-effectiveness threshold, r_c and g_k:

$$r_h \approx r_c - g_k$$

The intuition behind the model can be considered as follows. If $r_h < r_c$, a budget allocation that increases healthcare resources in the second period relative to the first will be preferred, permitting threshold growth. This unequal distribution of health spending between the two periods can be interpreted as saving from the first period to finance greater health spending in the second.

$$d_c = r_c$$
$$d_h \approx r_c - g_k$$

Consequently, the discount rate applied to costs should equal market interest rates, while the discount rate applied to effects should approximate market interest rates less threshold growth.

This formulation resembles differential discounting as advocated previously on the basis of a growing value of health over time, whereby the discount rate on health effects equals that of costs less growth in the value of health.[3]

Paulden and Claxton's Assumption

The key assumption of Paulden and Claxton's model is that healthcare resources are fully fungible over time; that is, they can be freely allocated between periods, subject to market interest rates, to maximise the present volume of health.

Reasons to Question this Assumption

There are two reasons to question the fungibility assumption: **(1)** it is unrepresentative of budget allocation practices in centrally-funded healthcare systems: healthcare systems tend not to systematically save or borrow between periods, but exhaust their annual allocations annually; **(2)** ongoing borrowing or saving is not sustainable beyond the two-period framework used by Paulden and Claxton. For example, continued threshold growth funded by saving from earlier periods to finance larger budgets in later periods cannot be maintained indefinitely.

Arguably, the fungibilty assumption may be justified in the short run. However, it is doubtful if the model reveals a long-term, sustainable relationship between market interest rates, threshold growth and the time preference for health.

Alternative Assumption for r_h

Rejecting Paulden and Claxton's conclusion that $r_h \approx r_c - g_k$ prompts the question of what discounting assumptions to use otherwise. The conventional assumption is that all goods have the same time preference, due to perfect fungibility of spending between alternatives and that this is approximated by market interest rates.[4] This implies $r_h = r_c$, which coupled with Claxton et al.'s conclusions yields:

$$d_h = r_c$$
$$d_c \approx r_c + g_k$$

This implies that the discount rate on costs should be adjusted upwards to account for threshold growth.

Consequences of Alternative Assumptions

Such alternative discount rates would be starkly in contrast to existing differential discounting practice, whereby the discount rate on health effects is adjusted downwards. For example, discount rates in the Netherlands used to be equal at 4%. Differential discounting was then adopted on the assumption that the growth in the value of health was 2.5%, resulting in discount rates of 4 and 1.5% for costs and health effects respectively.[5,6] Conversely, applying the adjustment described here and assuming for convenience that the threshold growth is also 2.5%, the resulting discount rates for the Netherlands would be 6.5 and 4% for costs and health effects respectively.

Accounting for threshold growth by adjusting d_c upwards rather than adjusting d_h downwards can have profound implications for cost-effectiveness ratios. For example, consider a preventative intervention with costs of €1,000 now, health gains of 0.1 QALYs and healthcare costs averted of €500 in ten years. The resulting present volumes of costs and effects and the cost-effectiveness ratios are shown in table 1 under equal discounting, differential discounting with the current Dutch rates and the alternative differential discount rates implied by the assumption that $r_h = r_c$.

Table 1: CE Ratios under alternative discount rates

	Equal discounting	Current Dutch rates	Alternative rates
d_c	4	4	6.5
d_e	4	1.5	4
Costs, €	649	649	716
Effects, QALYs	0.070	0.087	0.070
CE ratio, €/QALY	**9,200**	**7,400**	**10,200**

Adopting the alternative differential discounting results in the preventative intervention becoming less cost-effective with threshold growth. This result is intuitive, as no assumption is made about a change in the value of health over time, but the opportunity costs of future health spending averted has fallen due to threshold growth, so the future cost savings are now less important, furthermore, growth in the threshold has no bearing on the current costs of the intervention as they are incurred in the present.

Conclusion

Assuming that time preference rates for all goods must be equal may be a conservative assumption. A more thorough analysis would need to consider how the time preferences for health and other goods change with the factors that change the threshold, such as technology, healthcare demand and budgets. Paulden and Claxton's framework is a useful starting point for such an analysis. However, their assumption of healthcare resource fungibility is not representative of actual budget allocation practice. Consequently, further work is required to fully understand the correct form of differential discounting. CEA authorities such as NICE should wait until a robust consensus is achieved before recommending differential discounting.

References

1. Claxton K, Paulden M, Gravelle H, et al. Discounting and decision making in the economic evaluation of health care technologies. Health Econ 2011;20(1).
2. Paulden M, Claxton K. Budget allocation and the revealed social rate of time preference for health. Health Econ 2011; DOI: 10.1002/hec.1730.
3. Gravelle H and Smith D. Discounting for Health Effects in Cost-Benefit and Cost-Effectiveness Analysis. Health Econ 2001; 10(7).
4. Klok R, Brouwer WBF, Annemans LJP, et al. Towards a healthier discount procedure. Expert Rev. Pharmacoeconomics Outcomes Res 2005; 5(1).
5. Lipscomb J, Weinstein MC, Torrance GW. Time preference. In: Gold MR, Siegel JE, Russel LB, et al., editors. Cost-effectiveness in Health and Medicine. New York (NY): Oxford University Press, 1996.
6. College voor Zorgverzekeringen. Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie [in Dutch]. The Netherlands, 2006.

GLOSSARY

ACER	Average cost-effectiveness ratio. This represents the total costs of an intervention divided by its total effects.
BIA	Budget Impact Analysis. BIA is an analysis of the total net expected impact on healthcare budgets of an intervention.
CEA	Cost-effectiveness analysis. CEA is the appraisal of the costs and health effects of healthcare interventions.
CE Plane	The Cost-Effectiveness Plane is a graphical representation of the estimated costs and effects of interventions. Typically the do nothing strategy of no intervention is represented by the origin. Conventionally, the estimates of the costs and effects of alternative healthcare interventions relative to this reference strategy are represented on the horizontal and vertical axes.
Comparative Effectiveness	Comparative effectiveness is the comparison of the clinical effectiveness of alternative interventions. In contrast to cost-effectiveness, comparative effectiveness does not assess the resource implications of alternative strategies.
Cost-Effectiveness Threshold	The threshold is the upper limit on the ICER that is considered acceptable for reimbursement. Interventions with ICERs below the threshold are considered cost-effective while others are not. It can be considered society's limit on willingness to pay for care.
Discount Factor	A discount factor is a weight applied to costs and effects that occur in years other than in the discount year. Typically discount factors are applied to future costs and effects. They have weights less than one that fall over time, reflecting positive time preference, which is the tendency to prefer enjoying good things sooner rather than later.
Discount Rate	The discount rate is a measure of time preference. Positive discount rates reflect positive time preference, which is a desire to enjoy good things sooner rather than later.
Dominance	Interventions in the cost-effectiveness plane can be sorted into dominated and non-dominated strategies. Strategies that are more costly and less effective than other strategies are strongly dominated. Strategies that are more costly and less effective than a linear combination of two other strategies are weakly dominated. Those strategies that are not strongly or weakly dominated are the dominant strategies.

External Validity	This is the property that the results of the model adequately generalise to a real world setting. In the context of CEA this property is that the findings of a decision model correspond well to actual policy choices.
Efficient Frontier	The efficient frontier is the line that joins the set of dominant strategies in a cost-effectiveness plane.
HPA	Health Protection Agency. A UK government agency involved with the protection of public health.
HPV	Human Papillomavirus. HPV is a set of virus types that are a necessary cause of cervical cancer and known to be a causal agent of other gynaecological cancers, genital warts, penile and anal cancer, cancer of the head and neck recurrent respiratory papillomatosis.
ICER	The incremental cost-effectiveness ratio is the primary measure of efficiency employed in CEA. It measures the incremental costs of one intervention relative to another divided by the incremental effects between the two interventions. The ICER should only be estimated between non-dominated interventions that lie on the efficient frontier.
Net Benefit	Net benefit is a measure of cost-effectiveness expressed as either net monetary benefit or net health benefit, in which the benefits of an intervention are measured in terms of monetary units or health effects respectively. Net monetary benefit uses the cost-effectiveness threshold as a value of the health benefits produced by an intervention less the net costs of providing the intervention. Net health benefit measures the health benefit of the intervention less the opportunity cost of providing the intervention measured as the costs of the intervention divided by the cost-effectiveness threshold. Net benefit can be calculated over a range of threshold values.
PSA	Probabilistic Sensitivity Analysis is a method for assessing uncertainty in CEA. It involves the estimation of cost and effects over a range of possible parameter values, given the probability distributions of those parameter values and any correlation between parameters. The parameters are varied stochastically and the results estimated in many different iterations. The results are compiled to yield a joint probability distribution for costs and effects.