Tampere University

NEILL BOOTH

Economic evaluation, screening for prostate cancer, and 'value for money'?

Tampere University Dissertations 240

Tampere University Dissertations 240

NEILL BOOTH

Economic evaluation, screening for prostate cancer, and ‘value for money’?

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Social Sciences
of Tampere University,
for public discussion from auditorium F114
of the Arvo building, Arvo Ylpön katu 34, Tampere,
on 17 April 2020, at 12 o’clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Social Sciences
Finland

Responsible supervisor and Custos	Professor Pekka Rissanen Tampere University Finland	
Supervisor	Professor Anssi Auvinen Tampere University Finland	
Pre-examiners	Professor Hannu Valtonen University of Eastern Finland Finland	Professor Murray Krahn University of Toronto Canada
Opponent	Professor Colin Green University of Exeter England	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-1528-3 (print)
ISBN 978-952-03-1529-0 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1529-0

PunaMusta Oy – Yliopistopaino
Tampere 2020

To my maternal grandmother, for the wisdom of 'waste not want not', for a common-sense idea about economics, and for an ability to ask questions.

To my mother and father, who taught me to work hard and when to give up, respectively.

To Henri, for tenacity and verve, brightening my days on almost every occasion.

To Pirjo, for Sunday lunches and sundry advice: both have been a real life saver.

To Sam, for teaching me that opportunity costs don't rule in an uncertain world.

To Wend, for her continued support, especially when times were tough.

But most of all, to Anu, for waiting even longer than expected, and for the time, effort and love sacrificed to me and to this project, I hope there will be suitable opportunity to pay back a substantial debt.

Acknowledgements

From a first course in health economics by professor Richard Stevenson in Liverpool in the early 1990s, through sound teaching in economics by professor Martti Hirvonen in Tampere during the remainder of that decade, to Harri Sintonen giving extensive encouragement, bundles of support and pertinent documents, I have happily observed the second half of the life of health-economic evaluation. Without those three influences, I would not have set off on the track through health-economic evaluation which I eventually took. I also acknowledge the British and Finnish educational systems affording me instructive environments over the last 40 years or so; without this state-engineered foresight (or folly), my path would have likely been blocked at many turns. Financial support was gladly received, amongst others, from the Yrjö Jahnsson Foundation and the Doctoral Programs in Public Health.

My main thanks naturally go to my main supervisors: Pekka Rissanen and Anssi Auvinen, who in fairly equal measure, but in very different ways, have helped me build many of the foundations necessary for this work. Without their myriad of inputs to this undertaking, it would just not have been either possible, or so educational. I also profoundly thank the preliminary reviewers of earlier drafts of this study, emeritus professor Hannu Valtonen and professor Murray Krahn, for inspirational and highly valuable comments.

As this thesis mainly builds on what has gone before, I wish to recognize the contributions made by Galbraith, Quade and Ruskin, to name but a few, who did much to enlighten me about economics and economists. Steve Birch, Jo Coast and Cam Donaldson, are three more influencers who helped not only by publishing suitable material, but also by generously offering up their time, effort and wisdoms.

I wish to gratefully acknowledge half a dozen groups of people: 1) all the members during my time as part of the FinRSPC research group (especially, but not limited to, Henrikki, Irja, Johanna, Kimmo T., Kirsi, Liisa, Patrik, Paula, Tuomas K. and Ulf-Håkan); 2) research assistants who have helped this study along (Kirsi, Arja, Adela, Jani and last, but not least, Tomomi, with much appreciation for all their efforts); 3) Tampere Center for Child Health Research

-group members (especially professors Per Ashorn, Ulla Ashorn and Steve Vosti); 4) the members of the health-economics group in Tampere (again not limited to, Anna, Leena F., Leila, Olli, Satu, Suvi, Tanja, Terhi, Tiina J. and Tiina K.); 5) the staff of the Doctoral Programs in Public Health -network (especially Tiina, Kirsi and Marja); 6) all those academic information specialists who have helped in locating some of the more obscure literature reviewed for this study (in particular Sisko Kammonen and Pirjo Vuorio).

The following individuals can't escape a big thanks for their work, time or kindnesses: Teuvo L.J. Tammela for continued support throughout this long project, hopefully that investment will bring dividends!, Marjukka Mäkelä with erudite comments, often with the best of timing, and Timo Hakulinen encouragement and savvy behaviour throughout this process. Without the sagacity, exemplary action and good humour of Matti Hakama, the course of my Ph.D. would have been much different.

To Anna Shefl, Ph.D., Goddess of All Things Proofing, who has on many occasions provided a deal of fun, feedback and firm guidance (but who is not accountable for the language of this thesis). To the staff of Tampere School of Public Health (TSPH) in general, and to Raili Lepistö in particular, who made a stranger feel very welcome in a strange land. To Hannu Laurila for many a useful shove in what I think was the right direction, and for over two decades of continued support. To Steve Birch, a second mention, for doing a grand job in raising my game at least a little and doing much to help me finally drag my third Ph.D. article over the line.

To Jani Raitanen, for impressive patience and knowledge about, as well as willingness to share, statistical and other wisdoms. To Pasi Aronen, I know our academic paths have met more challenges than you might have liked, but I currently see them as full of value. Pasi, as a font of knowledge about matters economic and statistic, matters Helsinki, and things that matter, you have helped to make my journey much more palatable.

I would also like to express my thanks to colleagues during my times at TSPH, at the Finnish Institute for Health and Welfare (THL), at Peninsula College of Medicine & Dentistry and at Summaryx Ltd., for helping me on the road to an understanding of the craft of 'economic' evaluation. From my time at THL, I would especially like to thank Anneli Ahovuo-Saloranta for welcome normative guidance as to the nature of a classic Finnish Ph.D. thesis; Mikko Peltola, for buckets of positivity and help with the workings of register data; and Unto Häkkinen for a minor, but crucial, role. Thanks to all those in Exeter

in the summer of 2012, who gave me back a faith in health-economic evaluation, especially my learned host during that period. To those not already mentioned from Summaryx Ltd.: Iris, Jaana and Maija, I would like to offer my heartfelt thanks for your crucial support over many years.

Some very notable others who brought about meaningful changes during the course of this thesis include: Jussi Alitalo, Nigel Armstrong, Stina Björkell, Marja Blom, Phil Carmody, Owen Dempsey, Jack Dowie, Stiina Hänninen, Taru Haula, Ossi Hiekkanen, Timo Hujanen, Jean-Michel Josselin, Kirsi Karhu, Eila Kankaanpää, Vesa Kataja, Heli Koivisto, Perttu Lahtinen, Juhani Lehto, Pirjo Lehto, Ismo Linnosmaa, Teemu Murtola, Lien Nguyen, Leena Nikkari, Kirsi Norrbacka, Max Paddison, Anu Pakkanen, Henry Pearson, Markku Pekurinen, Helena Rantanen, Lars Sandman, Kimmo Savinainen, Heikki Saxén, Alex Scott-Samuel, Pirjo Siiki, Reijo Sund, and Heimo Syvälä.

In addition without the lovely, supportive and precious members of my family in England, I would likely not have finished this 'quest'.

Many of the above have acted as great guides, but it would be remiss of me not to name three mentors in particular – one from 'economics', one from 'statistics' and one from 'the medical community': The economist Eero Linnakko for sage advice on a host of subjects from costing weights through to flowers & bees, the mathematician Tapio Nummi for providing fun, fishy tales, amongst others, and the clinician Antti Jula for working so hard to support the early part of my research and for his calm and wholly exemplary style.

To the men who have provided responses to questionnaires during the course of this study, and to those who have allowed their register data to be used, I offer sincere thanks. To all others I have missed out or undervalued, please accept my apologies.

Finally, to Anu Planting, who saved me from a counterfactual that I can only gauge would have been more miserable, who I thank for filling me with energy and putting up with me when I thought times were hard, and even when I thought they were worse. As probably the best parts of my life, I also thank her for the invaluable gifts of Henri and Samuel.

Tampere, Midsummer 2019,

Neill Booth

Abstract

Economic evaluation has now been applied to health care for over 50 years, sometimes to good effect, sometimes for ill. This study seeks to give an understanding of what 'economic evaluation' can offer decision-makers, but also sets out to acknowledge its problems and pitfalls. In addition, this thesis applies one approach to economic evaluation, utilising data from the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC) after 20 years of the trial. Started in 1996, the FinRSPC is a pragmatic population-based trial investigating invitation to prostate-specific antigen (PSA) testing as a basis for mass screening to detect prostate cancer. Although the last invitations to screening were sent in 2007, health-care registers provide almost complete and almost continuous follow-up of *some* of the subsequent costs and effects, for up to 20 years in total.

This thesis includes four peer-reviewed articles related to economic evaluation or the FinRSPC, as well as a synthesis of those articles in a summary, which includes an integrated assessment of the development and current state of economic evaluation in the field of health care. The main rationale for much of this study is that care should be taken when interpreting economic evaluation. This is mainly because the language and terminology surrounding economic evaluation is often ambiguous and it often oversells what research can credibly offer. One main finding from this study is that there should be humility about our ability to adequately evaluate prostate-cancer screening. Data on costs and effects looks likely to remain deficient in some respects, due to problems in measuring outcomes, as well as due to problems in conducting trials on this topic. A second main finding is that economic approaches to evaluation inescapably include their own judgments about value; far from being a value-neutral form of assessment, these values determine its scope, content and appropriateness, to an important extent. Together with other findings, this study suggests a more classical approach to valuation would be useful: deliberative analysis. Such deliberative processes would include comprehensive quality assessment of information offered by any economic evaluation, integrating sources of information as and when necessary.

The articles which constitute this thesis each provide some evidence to support the suggestion of greater use of deliberative analysis. The measurement of health-related quality-of-life effects related to prostate-cancer screening in Article I faces challenges due to the timing and mode of data collection, due to the 'measures' used, as well as that much of the analysis is prone to biases. Article II mainly focuses on differences in average costs between the arms, as estimated from the accessible healthcare registers, for the men in the FinRSPC after 20 years. However, it should be noted that those estimates cannot account for, e.g., the potential impact of overdiagnosis or variations in treatment practices. Article III surveys a wide range of methodological literature and concludes that language and terminology related to health-economic evaluation should be used with sufficient humility to prevent it being oversold. It strongly recommends that the qualities of any 'economic' evidence be thoroughly checked for quality and relevance, especially if that evidence is to be utilised as part of priority setting processes. Article IV uses cost-effectiveness analysis to combine cost information from registers with analysis of data on effectiveness, in terms of different indicators of mortality, from the FinRSPC. The approach to economic evaluation used here revealed a (non-statistically significant) increase in mortality in the screening arm, but perhaps more importantly, it highlighted the uncertainties surrounding the evaluation of overdiagnosis associated with prostate-cancer mass screening.

In summary, although it is often claimed that economic evaluation can be useful in informing decision-making, in practice such claims are conditional on the qualities of the content of each economic evaluation. This thesis provides a rationale for assessing the qualities of all economic evaluations: their credibility should always be checked. Such quality assessment should look carefully at the uncertainties surrounding the evidence base, including any potential for medico-industrial influence. Every stakeholder should bear responsibility for appropriate interpretation of the available evidence, especially if interventions have consequences which are complicated, complex, or are simply not amenable to quantification or robust evaluation. Judging the appropriateness of the information from economic evaluation to any policy question should be approached with humility about what is, or can be, known.

This thesis concludes with three propositions. Firstly, that economic evaluation should continue almost unchanged, with an important exception being that the current overselling of economic evaluation, both in principle and

in practice, should be curbed. Secondly, attempts to improve the estimation of the likely costs and effects of interventions should continue, but with increased humility about the extent to which costs and effects can, in the foreseeable future, ever be 'measured' comprehensively. Finally, and perhaps most importantly, given that health-economic evaluation easily neglects many important impacts and, given the difficulties in changing language, terminology or research practices, those wishing to use information from economic evaluation should always engage in thorough critical assessment of its qualities.

Tiivistelmä

Taloudellista arviointia on sovellettu terveydenhuoltoon jo yli 50 vuoden ajan, joskus hyvin ja joskus huonoin tuloksin. Tämä väitöskirja pyrkii selittämään, mitä 'taloudellinen arviointi' voi tarjota päätöksentekijöille, huomioiden myös siihen liittyvät ongelmat ja karikot. Lisäksi väitöskirjassa sovelletaan yhtä taloudellisen arvioinnin lähestymistapaa FinRSPC-dataan tutkimuksen jatkuttua 20 vuoden ajan. FinRSPC (Finnish Randomized Study of Screening for Prostate Cancer) on vuonna 1996 alkanut pragmaattinen väestöpohjainen tutkimus, joka koskee miesten kutsumista PSA-testiin (prostataspesifinen antigeeni) eturauhassyövän systemaattisena joukkoseulontamuotona. Vaikka viimeiset seulontakutsut lähetettiin jo vuonna 2007, terveydenhuollon rekisterit mahdollistavat monien seulonnan aiheuttamien myöhempien kustannusten ja vaikutusten lähes katkeamattoman seurannan enimmillään 20 vuoden ajalta.

Tähän väitöskirjaan kuuluu neljä vertaisarvioitua FinRSPC-tutkimuksen taloudelliseen arviointiin liittyvää artikkelia, artikkelien synteesi yhteenvetona sekä integroitu arviointi terveydenhuollon taloudellisen arvioinnin kehityksestä ja nykytilasta. Väitöskirjan keskeisin lähtökohta on, että taloudelliset arvioinnit on syytä tulkita huolellisesti. Huolellisuutta tarvitaan, koska taloudellisessa arvioinnissa käytetty kieli on usein monitulkintaista ja sillä usein liioitellaan tutkimuksen mahdollisuuksia. Yksi tärkeä tulos tässä työssä onkin, että eturauhassyövän seulonnan arviointia varten saatavilla olevat tiedot voivat olla puutteellisia eikä tämä tilanne todennäköisesti ole muuttumassa. Tietojen puutteellisuus johtuu ongelmista sekä tulosten mittaamisessa että tutkimusten suorittamisessa. Toinen tärkeä löydös on, että taloudelliset arvioinnit väistämättä sisältävät omat arvoarvostelmansa. Taloudellinen arviointi on kaukana arvovapaasta tutkimuksesta, koska nämä sisäänrakennetut arvot määrittävät arviointien laajuuden, sisällön ja soveltuvuuden tarkoitukseensa. Yhdessä tämän tutkimuksen löydökset tukevat ajatusta, että klassinen harkitseva analyysi olisi arviointitapana hyödyllisempi. Harkitseviin prosesseihin tulisi sisältyä taloudellisen arvioinnin tarjoamien tietojen kattava laatuarviointi ja tarvittaessa lisätietojen huomioiminen.

Tämän väitöskirjan sisältämissä artikkeleissa esitetään näyttöä, joka tukee harkitsevan analyysin laajempaa käyttöä. Ensimmäisessä artikkelissa eturauhassyövän seulontaa koskevien terveyteen liittyvien elämänlaatuvaikutusten mittaaminen on haasteellista tietojen keräämisen ajankohtien, keräämistapojen ja kuvattujen mittarien vuoksi. Suuri osa analyysistä on lisäksi altis harhoille. Toinen artikkeli keskittyy lähinnä terveydenhuollon rekistereistä saataviin keskimääräisiin kustannuseroihin FinRSPC-tutkimukseen osallistuneiden miesten välillä 20 vuoden jälkeen. On kuitenkin huomattava, että näissä arvioissa ei voida huomioida mm. ylidiagnosoinnin tai vaihtelevien hoitokäytäntöjen mahdollisia vaikutuksia. Kolmannessa artikkelissa tarkastellaan monenlaista menetelmäkirjallisuutta ja todetaan, että terveystaloustieteellisten arviointien kielenkäytön ja terminologian tulisi olla riittävän asiallista, jotta vältyttäisiin liioittelulta. Artikkelissa suositellaan vahvasti, että 'taloudellisen' näytön laatu ja relevanssi tarkastettaisiin aina huolellisesti, erityisesti jos näyttöä käytetään priorisointiprosesseissa. Neljäs artikkeli hyödyntää kustannusvaikuttavuusanalyysiä yhdistäessään rekisterien kustannustietoja FinRSPC-tutkimuksen vaikuttavuusanalyysidataan hyödyntäen eri kuolleisuusmittareita. Tässä yhteydessä käytetty taloudellisen arvioinnin menetelmä paljasti (tilastollisesti merkityksettömän) lisäyksen seulontaryhmän kuolleisuudessa sekä toi esiin epävarmuustekijöitä, joita eturauhassyövän seulontaan liittyvän ylidiagnosoinnin arviointiin liittyy.

Yhteenvetona todetaan, että vaikka taloudellisen arvioinnin usein väitetään olevan hyödyllinen päätöksenteon apuväline, käytännössä hyöty riippuu kunkin yksittäisen arvioinnin sisällöstä ja kontekstista. Tämä väitös esittää vahvat perustelut kaikkien eturauhassyövän seulontaa koskevien taloudellisten arviointien piirteiden arvioinnille: niiden uskottavuus pitäisi tarkistaa kaikissa tapauksissa. Kaikkien sidosryhmien tulee kantaa vastuuta saatavilla olevan näytön tulkinnasta erityisesti silloin, kun interventiolla on monimutkaisia, komplisoituneita seurauksia tai näitä seurauksia on vaikea mitata. Taloudellisesta arvioinnista saatujen tietojen soveltuvuus päätösten perustaksi tulisi aina varmistaa ja pysyä nöyränä sen suhteen, mitä asiasta tiedetään tai edes voidaan tietää.

Väitöskirjassa päädytään esittämään kolme ehdotusta. Ensinnäkin, taloudellisia arviointeja ei ole syytä juurikaan muuttaa, mutta niihin liittyvää kielenkäyttöä tulisi tarkentaa ja liioiteltuja väitteitä vähentää. Toiseksi, työtä kustannusten ja vaikutusten arvioinnin parantamiseksi tulee jatkaa, pysyen nöyränä sen suhteen, missä laajuudessa kustannuksia ja vaikutuksia voidaan

kattavasti mitata. Viimeisenä ja kenties tärkeimpänä löydöksenä todetaan, että ottaen huomioon kuinka helppoa terveystaloustieteellisessä arvioinnissa on jättää monia tärkeitä vaikutuksia huomioimatta ja mitä vaikeuksia muuttuva terminologia, kielenkäyttö ja tutkimuskäytännöt aiheuttavat, terveystaloustieteelliseen arviointiin perustuvia tietoja käyttävien tahojen pitäisi aina arvioida huolellisesti kyseisten arviointien laatu.

Contents

List of Figures

List of Tables

Abbreviations

CEA	cost-effectiveness analysis
CERT	cost-effectiveness-ratio threshold
ERSPC	European Randomized study of Screening for Prostate Cancer
FinRSPC	Finnish Randomised study of Screening for Prostate Cancer
HEE	health-economic evaluation
HRQoL	health-related quality of life
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio
PCa	prostate cancer
PSA	prostate-specific antigen
QALY	quality-adjusted life year
RCT	randomised controlled trial
THL	Finnish Institute for Health and Welfare

Original publications

Article I: Booth N., Rissanen P., Tammela T. L. J., Määttänen L., Taari K. & Auvinen A. (2014). Health-related quality of life in the Finnish Trial of Screening for Prostate Cancer. European Urology, 65, 39-47.

Article II: Booth N., Rissanen P., Tammela T. L. J., Taari K., Talala K. & Auvinen A. (2018). Costs of screening for prostate cancer: Evidence from the Finnish Randomised Study of Screening for Prostate Cancer after 20-year follow-up using register data. European Journal of Cancer, 93, 108-118.

Article III: Booth N. (2019) On value frameworks and opportunity costs in health technology assessment. International Journal of Technology Assessment in Health Care, 35(5), 367-372.

Article IV: Booth N., Rissanen P., Tammela T. L. J., Kujala P., Stenman U.-H., Taari K., Talala K. & Auvinen A. (2019) Cost-effectiveness analysis of PSA-based mass screening: Evidence from a randomised controlled trial combined with register data. PLOS ONE, 14(11):e0224479.

1 Introduction

This introductory section sets out the some of the main reasons why this study, an article-based doctoral dissertation, aims to have two key roles. The first role of this study is to clarify what health-economic evaluation can mean in *principle*. The second role is one related to its potential role in informing decision-making through a specific application of health-economic evaluation to screening for prostate cancer in *practice*. With these dual roles, this thesis aims to provide sufficiently thorough guidance as to what economic evaluation can mean, at least in the context of its application to information relating to the health-economic evaluation of the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC).

There are three main reasons for proceeding in this way, the first is the originally planned content of this study, the second reason is the nature of the information which may be most suitable for the target audience, and the third reason relates to the apparent misuse and abuse of terminology in the field of health-economic evaluation.

The first reason for investigating both principles *and* practice is that this study will attempt, as far as possible, to follow the purpose set out in my application to begin undertaking doctoral research from the year 2002: "A cost-effectiveness analysis and an investigation into the long-term effects on health-related quality of life of screening will be undertaken as effectiveness and cost data is collected alongside the randomised FinRSPC. These investigations will then be combined with a description and assessment of the health-economic methodology used".

A second reason for the necessity of a description of both principles and practice is that the intended audience for the current study is *any* stakeholder involved in related decision-making processes. This conscious choice is made, firstly, from a belief that decision-making should ideally involve multiple stakeholders. Secondly, at the time of writing, the Finnish Ministry of Social Affairs and Health screening workgroup (16/10/2003–31/12/2007 and 1/9/2008–30/6/2015) has been dismantled, and screening-related appraisal-

processes appear to now be arranged on a case-by-case basis as part of the workings of the Council for Choices in Health Care in Finland (COHERE Finland). Given the current absence of any clear primary decision-making entity in Finland, this thesis aims towards providing, to decision-makers and the general public alike, information about the 'relevant' principles and the practical 'quality' of the evidence it presents. Therefore, this study aims to make both the principles and practice of health-economic evaluation more accessible and understandable by any stakeholder, regardless of their prior background knowledge about economic evaluation.

A third main reason for describing the principles which relate to health-economic evaluation in some detail is to avoid what can be seen as an overselling of health-economic evaluation through the ambiguity and corruption of terminology. Therefore, an integral part of ensuring that the results of this study can be more easily understood and interpreted correctly involves a clarification of some of the terminology used in and around health-economic evaluation.

As to the second role of this thesis, i.e., an application of health-economic evaluation to screening for prostate cancer in practice, here it will suffice to say that the original trial protocol for the FinRSPC included an assessment of cost-effectiveness as its secondary aim. The FinRSPC not only gave an opportunity to apply CEA using real-world data relating to a large cohort of men from a pragmatic population-based mass-screening trial, but, in conjunction with the relative flexibility of doctoral study in Finland, also gave wide scope as to the specific type of health-economic evaluation which could be undertaken. The articles presented as part of this thesis attempt to form a coherent whole, and one which is primarily aimed at promoting a deeper understanding of health-economic evaluation.

This thesis is organised as follows: Section 2 provides a background to the study. Section 3 briefly states the specific aims and formulates the general research questions for which answers will be sought in this study. Section 4 mainly reviews the methods applied, and techniques used, in principle. Section 5 turns to the practical application of those principles, as well as to briefly summarising the information this study adds to the literature. Some of the lessons learnt from this application health-economic evaluation, with accompanying suggestions and recommendations, are discussed in Section 6, with conclusions drawn in Section 7.

2 Background

As the focus of this study is, first and foremost, 'health-economic evaluation', the Background section sets out to define a working definition of what that can mean (Sub-section 2.1), and investigates both its development (Sub-section 2.2) and its conceptualisations (Sub-section 2.3).

2.1 A working definition of health-economic evaluation

The term health-economic evaluation has three intertwined parts: 'evaluation', 'economic' and 'health'. Although there seems to be no widely agreed definition of 'evaluation' (Stern, 2004), evaluation here refers to some form of assessment which assigns value, i.e., some process by which an estimate of, a calculation of, or statement of value is made or placed on an entity or intervention (Fox et al., 2016). Inherent in evaluative approaches are assumptions or statements about sources of value, often referred to as 'value judgements' (Neumann et al., 2016, p. 44). In the current study the normative, typically comparative, term 'evaluation' is qualified by the term 'economic'. 'Economic', here, is taken to mean any of the three main types of economics-based evaluative approaches (see, e.g., (Booth, 2019a)). Next, the term 'economic evaluation' needs *per se* to be qualified by the prefix 'health'. In this working definition, health refers to the topic or field to which the 'discipline' of 'economic evaluation' is applied (Culyer, 1981). Hence, health-economic evaluation (HEE) refers to applying some form of 'economic evaluation' to some topic related to 'health'. However, this is only a working definition, and it will immediately be suggested that the term is, in fact, too broad to be very useful, unless it is specified more fully in practice. This is because different approaches to economics define value in such disparate ways. As an alternative, simply referring to 'economic evaluation', as per the title of this study, is intended to highlight a need to understand the vagaries of gauging 'benefits' and 'costs', i.e., of placing a 'societal' value on an entity or intervention (see, e.g., (Hammond, 1958) and (Wildavsky, 1993)). An attempt to defend these claims will form a central part of this study.

2.2 Understanding the history of (health-)economic evaluation

Although we now have a working definition for health-economic evaluation, here the view is taken that it is only possible to properly understand health-economic evaluation when we are informed about its history and development (in accordance with, e.g., (Keynes, 1924) or (Hodgson, 2001)). Of course a caveat is required here, a variety of histories can be written, each with its own legitimacy, and none will be wholly adequate (Backhouse, 2005). However, this thesis contends it is important to reflect not only on the most recent trends and developments in health-economic evaluation, but also to put those into perspective. Current health-economic evaluation may be understood better with an appreciation of the challenges 'economic evaluation' has already faced.

Different types of evaluative research, and different ways of informing policy and planning, have likely been central components in the development of society throughout recorded history (Trefethen, 1954). Groups such as civil servants, academics, and private analysts have long provided information, evidence or advice to government planners, decision-makers or societies on a variety of policy-related, i.e., politico-economic, problems. Although the history of the use of 'cost-benefit thinking' dates back at least a millennium (see, e.g., (Quade, 1971b) or (Campbell & Pryce, 2003, p. 30)), one of the first printed records relating to 'political economy' is only just over 400 years old (de Mayerne Turquet, 1611). One way to interpret such early examples of political economy is the "specific advice given ... to governments or to the public at large either on broad policy issues or on particular proposals" (Mishan, 1982a, p. 13). It is in this sense, as the application of economic principles to practical research in order to provide information for decision-making (Bentham, 1843), which the 'economics' or 'political economy' of policy analysis will be understood here (see, e.g., (Meade & Hitch, 1938, p. 221) or (Wildavsky, 1993, p. 122-123).

From the perspective of economic-evaluation frameworks, two of the most important concepts within political economy received much impetus in their development during the European Enlightenment of the 18th Century, namely, the related evaluative notions of both '**cost-benefit thinking**' (see, e.g., (Steuart, 1769, pp. 317–345), (Franklin, 1772), and (Bentham, 1843) and '**opportunity cost**' (see, e.g., (Franklin, 1748, pp. 375–377), (Cantillon, 1755) & (Smith, 1776: Book I, Chapter VI, p. 1)). Closely related to both of these

ideas is a third concept which came into wider use only after the start of the 19th Century, namely, '**efficiency**'. It is important to note that here the term efficiency does not refer to the earlier philosophical or religious usage, in the sense of fitness or power to accomplish, or success in accomplishing, the purpose intended (Browne, 1658). Instead, the main meaning of efficiency in this current study can be found from its 'technical' usage, when employed by engineers and physicists to express or measure the performance or productive capacity of machines, i.e., a means of assessment of the ability of something to achieve *specified goals* (Alexander, 2008, p. 2; Franklin, 1748). The interrelated notions of 'cost-benefit thinking', efficiency and opportunity cost form an evaluative triad of particular importance to valuation in general, as well as to health-economic evaluation in particular. It is also useful to have a working definition for 'opportunity cost', here. Central to almost all definitions of opportunity cost is the idea that it represents "the opportunities foregone in accepting a certain line of action" (Green, 1894). However, it is important to note that alternative approaches to health-economic evaluation tend to place very different emphasis as to what constitutes opportunity cost, i.e., alternative valuation frameworks within economic evaluation vary in their approaches to the valuation of "opportunities forgone" (Article III, (Booth, 2019a)).

Of course, it is necessary to operationalise the above evaluative triad in practice, too; otherwise these notions remain little more than empty concepts or abstract principles which *in practice* can offer anything from "an infallible means of reaching the new Utopia to a waste of resources in measuring the unmeasurable" (Prest & Turvey, 1965). One set of approaches to the operationalisation of the notions of cost-benefit thinking, efficiency and opportunity cost, is through the formalisation (of one notion) of 'value' via neoclassical economics and subsequent developments in welfare economics (Mazzucato, 2018; Mirowski, 1990). However, it should be noted that this set of approaches would come to be used mainly in those health-economic evaluations based on the aggregated 'utility' of affected individuals, i.e., in welfarist approaches to economic evaluation. An alternative, second route attempting to address the evaluative triad was developed, which has been referred to as both 'non-welfarist' and 'extra-welfarist', which involves the possibility of overruling the above individual judgements of 'utility' (Culyer, 1989). In this way the 'extra-welfarist' approach fundamentally diverges from, neoclassical economics in general, and from Paretian welfare economics

in particular, by focussing on some notion of 'societal utility', instead of the welfarist notion of 'individual utility'. It should be noted here that the majority of health-economic evaluations would come to follow the extra-welfarist approach in the first 50 years of the regular application of economic evaluation to the field of health. Although early examples of cost-effectiveness analysis can be found (Priestley, 1831; Rumford, 1876; Editorial, 1899), more formal mathematical approaches would have to wait until operations research became popular (see, e.g., (Dantzig, 1951, pp. 339—347) and (Smith Jr et al., 1953)). Important elements in the rise of popularity of operations research and related methods include their application to military planning after the second world war (Specht, 1960); the large-scale incursion of mathematicians and physicists, in particular, into neoclassical economics (Mirowski, 1999); the promotion of a doctrine of 'rationality' (Colvin, 1987, p. 13); as well as that its results seemed to promise both stability in, and the ranking of, input-output relationships (Alexander, 2008, p. 64). It should be remembered that the utilisation of cost-benefit thinking occurs via a multitude of different forms of efficiency analysis (under a variety of names including operations research; systems analysis; cost-benefit analysis; cost-effectiveness analysis; and cost-utility analysis (Fisher, 1965)), radically expanded in the early 1960s (see, e.g., (Quade, 1971b)). However, approaches to cost-benefit and systems -thinking became somewhat less popular in some of its applications, as the 1960s progressed (Worthley, 1974). Its limited applicability to certain topics is one reason for its fall from popularity (Quade, 1971b), e.g., when straying away from its traditional roots in application to the technical realm (Godin, 2007). Moving into the social realm brought with it difficulties such as identifying an acceptable goal or goals, as well as difficulties in measuring to what extent the chosen goals were being achieved (see, e.g., (Hitch, 1967) or (Rickover, 1967)). Subsequently, in addition to the differing formal approaches to operationalising cost-benefit thinking outlined above, a more nuanced understanding of the pitfalls and potentials of efficiency measures had been achieved (Quade, 1970; Schultze, 1968).

Indeed, for some, the application of more informal cost-benefit thinking, with its accompanying judgement and deliberation, have retained an important role to this day (Daniels & van der Wilt, 2016). Such nuanced understanding and concern for case-by-case assessment of the qualities of

health-economic evaluation is referred to below, and in Article III, as deliberative analysis (Booth, 2019a).

As set out below, in Sub-section 6.2.3, deliberative analysis will include comprehensive quality assessment of health-economic information, integrating other sources of information as and when necessary. Wherein both the information underlying any CEA, as well as the CEA itself, should be checked for its credibility and appropriateness; judgment; thought and expertise will likely be an advantage (Culyer, 2015; Dror, 1970). As a brief summary of this section, it is useful to note that different forms of CEA-related analysis have long provided information to politico-economic, planning and policy processes, and has the ability to do so without jeopardising the role of analysts as analysts (Quade, 1971a, p. 295), rather than as advisers or mongers.

2.3 Conceptualising health-economic evaluation

Health-economic evaluation (HEE) can be classed as either an applied social science (Williams, 1989), or, alternatively, simply as an example of the application of social research or evaluation (Sculpher et al., 2005). For this reason alone conceptualising HEE is challenging: it covers a multitude of ideas. Regardless of which label is used, such research is generally conceptualised as an attempt to *inform* the type of problem which asks: "What should we do?". In line with a Knightian view of economics (see, e.g., (Knight, 1951) or (Hicks, 1986)), the claim will first be made here that many of the concepts of neoclassical economics are inherently restricted in their applicability to applied policy analysis, such as that undertaken in extra-welfarist HEE (Klein, 1989). This is largely due to the imperfections of the market for health care, such as asymmetry, or lack, of information about health-care (Arrow, 1963). Such a restriction would suggest extra-welfarist HEE (EWHEE) would lack any clear legitimacy for a prescriptive role in decision-making processes and, moreover, suggest EWHEE is merely applied research (Sculpher et al., 2005).

Welfarist forms of health-economic evaluation (WHEE) of course do include elements of neoclassical economics (Birch & Donaldson, 1987), but it is argued here that grounding health-economic evaluation firmly in conventional welfare economics would overlook the numerous objections to its practical use: whether in the form of orthodox cost-benefit analysis

(CBA), contingent valuation, or in the form of mathematical programming. Objections include the refuted separation of inquiries into economic welfare from inquiries into freedom, rights, equality and justice (Hausman et al., 2017), the market-mimicking tendencies of these approaches (Sandel, 2013), in addition to the problems of measuring individuals' utility and the informational requirements of a mathematical-programming approach (Drummond et al., 2015, pp. 116—118).

Indeed, rather than being classed as a decision-*making* tool, it would seem more appropriate to suggest WHEE and EWHEE merely produce information. Further, it is suggested here that this information can inform, or misinform, decision-making processes. Suggesting this rather limited role as a source of information restriction seems appropriate given the inherent dependence of any research on the definition of the relevant 'ends', i.e., on the need for the definition of the objective(s) or goal(s) of policy (Knight, 1956, p. 134). According to this viewpoint, health-economic evaluation has little reason to claim that it provides a solution to decision-making problems (Ashmore et al., 1989). The response from proponents of EWHEE is typically that the 'solution' they offer is only one of a number of possible solutions to priority-setting problems (Williams, 1991). Nevertheless, proponents, of EWHEE in particular, often use a lexicon which includes concepts framed in terms of 'decision rules', 'solutions' and 'efficiency' (see, e.g., (Karlsson & Johannesson, 1996) or (Weinstein, 2012)), even if many authors do not typically recommended strict adherence to such 'rules' (see, e.g., (Drummond, 1980), (Lord et al., 2004) or (Sculpher et al., 2005)). For example, the term 'efficiency': originally restricted to physics and engineering, may be much less appropriate when applied to matters of political economy (de Santos, 2009). Efficiency is not only a problematic label, in order to achieve successful conceptualisation or operationalization it demands careful measurement and determinable input-output relationships, a point which will be returned to in Sub-section 4.4. below. These terms not only have a specific meaning within HEE, but can also be misinterpreted by the laity who may not be aware of their specific and technical nature when they are similar to terms used in contexts unrelated to HEE (Meade & Hitch, 1938, p. 221). One key point which can be drawn here is that technical solutions related to 'efficiency' have been borrowed from the natural sciences and then, largely unchanged, have been reused, i.e., reused with their nomenclature largely intact (Knight, 1923). A second key

point is that a 'market-mimicking' element of EWHEE (see, e.g., the library-bookshelf metaphor in (Culyer, 2016a)), is a construct which drives the potentially misleading notion that 'use' of EWHEE for public policy choices would necessarily have considerable political significance (Culyer, 2016b). Such an apparent lack of humility seems even more strange given that HEE had already been practiced for many years (Drew, 1967) before quasi-theoretical justifications for EWHEE appeared (see, e.g., (Culyer, 1989) and (Garber & Phelps, 1997) or (Meltzer, 1997)). There would also seem to be a trade-off to be faced: if effort is expended on making data and assumptions fit into the logical and 'empirical' constraints of some theoretical EWHEE framework, less effort will be available for ensuring that other informational needs are served.

Practitioners of EWHEE, for example, can, of course, claim CEA is akin to the 'utility maximization' underlying WHEE. However, doing so they would need to accept some fairly heroic axioms relating to expected utility theory and 'presumed' Pareto efficiency. For many commentators, these axioms are violated too often to be an acceptable rationale for using CEA for 'utility maximization' (McGuire, 2001, p. 10; Sculpher et al., 2005, p. 10). The view taken here follows Blaug, in that it assumes policy debates will typically turn into fuzzy comparisons between slightly incommensurate entities, and that any dependence on welfare economics, or for that matter WHEE, would be unwarranted (Blaug, 2007).

As an alternative, two main strategies for undertaking CEA outside the constraints of orthodox welfare economics have flourished in the literature (see, e.g., (Drummond et al., 2015, pp. 267—339)). One is more reliant on the aforementioned extra-welfarist theorising and has strong foundations in decision science (see, e.g., (Weinstein & Zeckhauser, 1973) and (Claxton, 1999)), and a second is not strictly bound to any economic theory *per se*, but is still faithful to the concept of efficiency and can still be highly pertinent when informing the decision-making process (see, e.g., (Quade, 1971a, p. 1) and (Sugden & Williams, 1978, p. 191)). The strategy which is largely based on decision science, with its reliance on the concept of expected value (Arrow & Lind, 1970), has been challenged in the literature (see, e.g., (McKean & Moore, 1972) and (Mishan, 1972)). Decision-analytic approaches have also been subjected to criticism with respect to a number of assumptions (Krahn et al., 1997). Firstly, the assumption that sufficiently clear and consistent input-output relationships exist seems unlikely to hold in applications to

health care (see, e.g., (Kothari & Birch, 1998), (Kristiansen & Mooney, 2004) and (Lessard & Birch, 2010)). Secondly, in respect of the assumption that efficiency is, *per se*, a desirable end. The desirability of the chosen efficiency indicator can be criticised if the relevance to the decision-maker of that indicator is diminished, either due to the content of the indicator or due to flaws in measurement or analysis of that indicator (see, e.g., (Mugford, 1989), (Kristiansen & Gyrd-Hansen, 2006), (Alexander, 2009, p. 1019) and (Adami et al., 2018)). Such measurement flaws can be used to critique the 'theory' underlying EWHEE, given its reliance on a price-per-effectiveness unit being determinable, determined and an appropriate measure for systematic comparison (Culyer, 1989). Another important reason to question the EWHEE approach to defining 'value' is that its definition of value seems unduly based on neoclassical market-clearing notions; in short, price connotes value, (i.e., value is assumed to relate to price, e.g., 'price per QALY', (Claxton et al., 2008)). Of course, as already noted above, 'markets' for health care typically do not function smoothly, amongst other things, due to severely impacted and asymmetric information; imperfect agency relationships; inherently monopolistic tendencies among suppliers; absence of prices for all entities; substantial scope for both moral hazard and adverse selection; and a host of issues related to a multitude of equity considerations; as well as due to related to possible externalities, spillovers or higher-order effects (Culyer, 2008, p. 60). However, even in the absence of a functioning market, many ardent extra-welfarist health economists typically choose to define value through such a 'price'. This also means EWHEE requires the creation of something that resembles a 'market-clearing price' via the location of, or the fabrication of, an efficiency threshold (Caro, 2009). In other words 'fixing' a threshold price, a pseudo-price, in terms of some form of efficiency, for any technology being assessed. Applying a standard EWHEE approach is avoided in the current study, in part because such reductionist marketization of would *de facto* be antithetic to a strict interpretation of health-economic evaluation being able to *inform* decision makers about *allocation* (Lübbe, 2011, p. 108). Indeed, although statements about such a 'market-clearing solution' may have meaning within academic extra-welfarist health-economic evaluation, justification for adherence to them in any wider societal discourse is much less apparent. Insistence on economic-textbook notions of the use of the EWHEE (Culyer, 2016a), would seem to be highly arrogant, given the absence of some aspects of morality

from much of EWHEE (Smith, 1761). Further reasons for treating CEA based on the EWHEE approach with a great deal of scepticism will be set out in Sub-section 6.1. and will include practical problems associated with modelling, e.g., practical issues which arise due to problems with measurement, inadequate data, the assumptions often required to underpin EWHEE, or any combination of these.

A second form of CEA does not necessarily resort to the above EWHEE framework, but instead can rest on the simple concept of a comparison of alternative courses of action in terms of their costs and their effectiveness in attaining a specified objective (Quade, 1971a, p. 294). When using this approach, comparisons are often based on the available resource use and outcome data, collected via, e.g., health services research, on all, or on a subset of, trial participants, often in the form of an analysis of patient-level data from randomised, controlled trials (RCTs) (O'Brien, 1996). Just as Drummond et al. eschew references to microeconomics in their latest textbook (Drummond et al., 2015), the single-study based approach to CEA used here is not necessarily reliant on any theoretical microeconomic framework (Ramsey et al., 2015). When free of microeconomic postulates, the appropriateness of the evaluation will largely depend on the qualities of the RCT in question (Lessard & Birch, 2010). In this second form of CEA, as knowledge from the discipline of economics no longer needs be combined with knowledge from some other discipline, the problem of 'discipline mix' may be largely avoided (Valtonen, 1993). One point to note here is that positivist clinical epidemiology and decision sciences seem to dominate much of applied HEE, with methods in HEE typically no longer being those of economics *per se* (Birch & Gafni, 2004, p. 55; Lessard & Birch, 2010; Sheldon, 2005). In CEA in particular, although some of the items to be included in the evaluation can be given an economic foundation, CEA *per se* does not rely on economics, of course, economics is sometimes used to offer guidance when conducting a CEA, but economics does not tend to dictate the economic evaluation process *per se* (Neumann et al., 2018). To be clear, although quasi-economic theory can be fitted to EWHEE approaches, it is only WHEE approaches which ensure adherence to orthodox economic principles, and then only to welfare-economic principles. More will be said about the limitations of positivist approaches to evaluation below.

In addition, although it is now widely accepted that the assumption of 'utility'-maximising CEA is almost impossible to justify within the health

sector (Hurley, 1998), as noted above some heroic assumptions would also be needed to justify the assumption of 'health'-maximising CEA. This is because, *in practice*, the measure of effectiveness, i.e., some measure of health impact, would need to be acceptable to society as a whole (Valtonen, 1993, p. 195). In Knightian terms, unless society can agree on the 'end', i.e., unless society can agree on a single desired measure of effectiveness, CEA will not maximise 'utility' or 'health' (Hammond, 1991). Instead, any single CEA will only inform us what can, given the assumptions, data and analysis used, be classed as 'efficient', in terms of the particular measure(s) of effectiveness used. This interpretation of CEA as a form of efficiency analysis brings with it its own potentials and problems, and how these interact with the concept of opportunity cost will be returned to below.

The two main points from this second section are, firstly, that the concepts of cost-benefit thinking and efficiency existed long before they became cloaked in any form of formalised economics or comparative analysis. It is primarily for this reason that the current thesis does not employ or describe formal theory related to CBA, CEA or health-economic evaluation, but instead, promotes a pragmatic appreciation of both CEA and cost-benefit thinking. Secondly, HEE is too ambiguous a concept to be categorised with any brevity, if at all, it will therefore be necessary to further investigate the idea in Section 4.

3 Study questions and objectives

The principle aim of this study is the presentation of understandable information relating to 'economic' evaluation and prostate-specific antigen screening for prostate cancer from the FinRSPC. In line with these aims, this study sets out to achieve two central objectives: 1) increasing understanding about what economic evaluation can mean, as well as 2) increasing understanding about the results from the application of CEA to the FinRSPC after 20 years of the trial. These two objectives are achieved via the three research questions described below, one focussing mainly on principles relating to economic evaluation, a second relating mainly to the practice of economic evaluation, as well as a third, turning to how the information from this study, and similar studies, might be best interpreted.

1) How should (health-)economic evaluation be understood, and what are some of the ways it can be misunderstood?

2) What can be known about the costs, effects and cost-effectiveness of prostate-cancer screening after 20 years of the FinRSPC?

3) What are some of the problems, pitfalls and potentials of the application of economic evaluation to a topic such as screening for prostate cancer?

Research question **1)** is addressed by Article III and by Section 4. Research question **2)** is addressed by Articles I; II & IV, as well as by Section 5. Research question **3)** is addressed to some extent in all articles, but is also discussed in Section 6. Together, the three research questions above also have the objective of providing a response to a broader practical question: "Can economic evaluation help us to know if prostate-cancer screening is 'worthwhile'?". However, although some discussion of this broader (policy) question will run throughout Section 6 of this study, it should be noted at the outset that the answers to such questions should be treated as preliminary, as they transcend 'scientific' or research evidence. It has long been acknowledged that applied social research cannot typically provide answers to questions of value, instead the role for a study such as this is to provide evidence which stakeholders can then deliberate upon (Weinberg, 1972).

4 Applying 'economic evaluation' to the FinRSPC: Some principles

After describing the broad principles of health-economic evaluation above, but before setting out the way in which health-economic evaluation is applied here, it will be useful to describe cost-effectiveness analysis (CEA) in a little more detail. Although early examples of crude CEA in health care can be found (see, e.g., (Editorial, 1899)), this brief review attempts to explain what CEA typically means today, by reference to both its wider use in decision-making and planning during the last century (see, e.g., (Hitch, 1953)), as well as to its more recent history. This section touches upon some of the content of Article III, but the main focus here is on the problems and potentials of applying 'economic evaluation' to the FinRSPC. The interested reader can, e.g., in Article III, find more detailed explanation and critique of the different 'economic' standpoints from which economic evaluation can be undertaken.

4.1 Cost-effectiveness analysis (CEA) in principle

Cost-effectiveness analysis evaluates a ratio of some indicator of average effectiveness for two (or more) interventions relative to some indicator of the (mean) costs associated with those interventions. In simple terms CEA calculates or, rather, typically estimates, the incremental cost-effectiveness ratio (ICER) using the following generic formula:

$$ICER = \frac{\Delta \mathrm{C}}{\Delta \mathrm{E}} = \frac{\mathrm{C}_{intervention} - \mathrm{C}_{control}}{\mathrm{E}_{intervention} - \mathrm{E}_{control}},$$ where $\mathbf{C}$ is some measure of average cost and $\mathbf{E}$ is some measure of average effectiveness.

The numerator in the above formula is an estimate of the difference in some of the costs between the two alternatives and is divided by the denominator, i.e., an estimate of the difference in one, or some, of the health-related outcomes. In the rightmost part of the equation above, two alternatives are denoted by their subscripts, first, an intervention (e.g., in the current study it

will refer to the screening arm of the FinRSPC) and, second, a control or reference group (which refers to the control arm of the FinRSPC in the current study). In principle, the CEA ratio can be used to help maximize the estimated units of effectiveness produced from a given level of resources or, alternatively, minimise resource use for a given estimated number of units of effectiveness (Gafni & Birch, 2006).

4.2 A wider review of health-economic evaluation

In many industrialised countries healthcare has, since the end of World War II, and especially since the end of the cold war, received increasing amounts of support and investment from government. Just as the 'military-industrial complex' before it, this 'medical-industrial complex' has generally included arrangements for the evaluation of proposed investments (see, e.g., (Steiner, 1965), (Butterfield, 1968), (Pole, 1972) and (Relman, 1980)). However, the extent to which industry is involved in current governance processes in health care in any jurisdiction, and the extent to which there is regulatory capture, remains difficult to ascertain with any certainty (Banta, 2018; Berry, 2017, pp. 199—214). What is more certain is that the 1960s brought with it an increased interest in what analysis and evidence could offer, or could be offered to, planning authorities (Specht, 1960). The level and type of research to be undertaken was closely intertwined with what the state deems necessary for underwriting the technology of modern industrial enterprise (Galbraith, 1979). Such state involvement was exemplified in the US with the establishment there of both the Planning, Programming, and Budgeting (PPB) -system and 'technology assessment' itself (see, e.g., (Schick, 1966) and (Daddario, 1967)). A number of other countries gradually followed this US lead, for example, the principles relating to the application of CEA to health care were set out in the UK by the early 1970s (see, e.g., (Walsh & Williams, 1969), (Culyer et al., 1971) and (Cochrane, 1972 (2004))). However, it would be over two decades before a health technology assessment (HTA) programme was set up as part of the first explicit research and development strategy of the (then UK) National Health Service (NHS) (Department of Health, 1993). On the basis of the review of the literature undertaken for the current study, it is claimed here that, combined with other influences (e.g., (Torrance, 1970: p. 44)), much of the practical application of health-economic evaluation was largely pushed forward, by the

more general advancement of CEA, the PPB, and the Office of Technology Assessment in the US (see, e.g., (Office of Technology Assessment, 1976) and (Office of Technology Assessment, 1980)), with some of those early influences mirrored by more specific development of CEA in the health sector in the UK (see, e.g., (Williams, 1967), (Drummond, 1978) and (Sugden & Williams, 1978)).

4.3 How is CEA related to cost-benefit thinking?

At present, the most prominent approach to health-economic evaluation and CEA, which was largely developed in York, by Williams, Culyer and Drummond, among others, is extra-welfarist. This approach is centred around the measurement of 'health-related quality of life' (HRQoL), often with durations in HRQoL states as the chosen indicator of medical output, i.e., as the denominator in the ICER (see, e.g., (Culyer et al., 1971), (Pole, 1974) and (Culyer, 1989)). As already noted above, such 'non-welfarist' approaches diverged from neoclassical welfare-economic traditions, which were deemed to have outlived their usefulness (Williams, 1976). Therefore, CEA was recommended as a more practical form of cost-benefit thinking (Drummond, 1981) or, to put it in the blunt words of Mishan: "CEA is a truncated form of CBA" (Mishan, 1988, p. 110). The important point here is that, by moving away from an 'all-encompassing' welfarist approach of cost-benefit analysis or analysis through mathematical programming, EWHEE would develop into a rather disordered set of approaches only loosely related either to economics, or to each other (Gerard, 1992). Despite efforts to standardise EWHEE through guidelines and the issuing of articles on best-practice (see, e.g., (The Professional Society for Health Economics and Outcomes Research (ISPOR))), it remains a rather diverse set of approaches in practice because it *could* accept any maximand: the maximand may be whatever the analyst **or** policy-maker deems appropriate (Culyer, 2014, p. 203). It should be carefully noted that in the absence of a decision-maker with a known objective or maximand, the analyst may suggest and use their favoured maximand. More worryingly perhaps, even in the presence of a maximand or maximands chosen by a decision-maker, analysts may for the sake of convention or convenience still supplant their own indicator to approximate that maximand. The flexibility in scope of any specific instance of economic evaluation brings with it judgement and decision (Culyer, 2008), i.e., it requires value judgements rather than

necessarily being an objective choice in advance (Hitch, 1971; McIntosh et al., 1999). Although one maximand is customary in HEE, the most appropriate maximand(s) will depend on specific contexts of culture, history and political aspiration (Culyer, 2008; Mishan, 1982b). However, if HEE attempts to serve the information needs of planning problems, a single maximand will rarely suffice. Indeed, truncation in scope means that EWHEE serves only *some* of the information needs of cost-benefit thinking. To fully meet the needs of broad evaluation the chosen maximand in CEA would be required to be 'all encompassing', i.e., the chosen indicator or measure would need to encompass the value which should be attached to programme impacts (Klarman, 1982, p. 594). CEA in general, and EWHEE in particular, rarely succeed in producing a broad evaluation, and to presume otherwise can be seen as a mere illusion (Wolff, 2004).

Given the above divergence between the information provided by CEA and the information which would ideally be available for social planning problems, two main approaches have been followed (Orr & Wolff, 2015). The first approach could be described as humble in its ambitions, with the second seeming somewhat more arrogant. The first is that information from CEA is presented without extensive interpretation through a theoretical framework, ICER-related information stands on its own merits, leaving the users of that information to interpret it in a manner in line with their values. The second approach, the approach typically taken under EWHEE, is to use some threshold to create or mimic a market (Galbraith, 1973; McCabe et al., 2008). Such thresholds, i.e., efficiency standards, are compared to ICER estimates, and can immediately take on a (non-Paretian) allocative aura (Gyrd-Hansen, 2014). The result of an EWHEE, an estimate of economic efficiency in the form of a cost-effectiveness ratio, is typically only a partial indicator of the term efficiency as it used in every-day language. Thus the ICER is a partial indicator, an indicator of economic- or pseudo- efficiency (Green & Barker, 1988). Despite this some have claimed that EWHEE approaches approximate an interventions value (Garrison et al., 2018), phrases akin to 'the ICER shows whether or not the intervention represents value for money' are used in place of stricter language 'the ICER estimate is an indicator of the efficiency of the intervention, albeit incomplete' (see, e.g., (Maynard & Bloor, 1997), (Moore et al., 2010) or (National Institute for Health and Care Excellence, 2019)). In engineering, almost all relevant aspects of efficiency can be quantified in an accurate, if not precise manner. In contrast, especially

in the politico-economic realm, many dimensions of importance will lie outside CEA-based indicators of efficiency (Alexander, 2009, p. 1011). In the field of orthodox cost-benefit analysis it may be appropriate to talk about 'value for money', e.g., 'optimising net social costs and benefits' (Lowe, 2008), but does not seem so appropriate in the field of health-economic evaluation (Birch & Gafni, 2006), or in the field of medicine (Gusmano & Callahan, 2011).

Despite the concerns outlined above, CEAs serving extra-welfarist goals, using QALY-based approaches, have gained a dominant position as the most popular form of health-economic evaluation in practice (McIntosh et al., 2010, p. 2). Nevertheless, it is important to note that all approaches to CEA can also serve more 'classical' cost-benefit thinking, a point which will be returned to below (see also Article III for further details).

4.4 The (almost) central role of efficiency?

Although economic efficiency, in its neoclassical economic form, has been seen as having a primary role in health-economic evaluation, it will not always have primacy when planning resource allocation (see, e.g., (Maynard, 1999), (Musgrove, 1999) and (Robinson, 1999)). Of course, as a 'measure' of efficiency, CEA can be an *indicator* of value, but CEA is typically merely only an indicator of an already truncated or restricted notion of value, rather than necessarily providing a broad evaluation (see, e.g., (Quade, 1971a), (Konu et al., 2009), (Häkkinen, 2011) or (Hämäläinen et al., 2019)). The ability of CEA to produce relevant information will depend both on what goals are being assessed and on the responsiveness of each effectiveness metric used (Porta, 2014; Waddington, 1986).

The type of efficiency which is estimated by EWHEE or WHEE receives much of its meaning in relation to the achievement of goals (often simplified, measurable or pre-specified). Nuanced 'higher-level efficiency', i.e., societal value, must embody some procedure to include the respective goals of all stakeholders, to some extent (Wiseman, 1989b). To aim uncritically for efficiency has been seen by many as a mistake: most people would need look no further than highly-efficient National Socialist German Workers' Party's extermination camps to realise that (Culyer, 2015). It follows that any intrinsic goodness in any 'efficiency' metric depends on elements of the value

we attach to the goal(s) we are being efficient at (Churchman et al., 1954). Efficiency metrics cover, to a greater or lesser extent, relevant elements of value (Culyer, 2015). To the extent that some element of value which is considered to be important is not included in the effectiveness or efficiency metric, it fails as a valuation mechanism (see, e.g., (Ruskin, 1872, pp. ix-xiv) and (Knight, 1923)). In every case, if the operationalised 'measure', metric or indicator does not adequately measure what we hope it to measure, it will also fail as a comprehensive valuation mechanism (Joint Economic Committee, 1967). Although, on the one hand, cost-per-QALY efficiency metrics offer one popular approach to economic evaluation, on the other hand the policy context is also seen as central in determining what is the appropriate methodology to use (Tsuchiya & Williams, 2001, p. 43). One problem in choosing the most appropriate efficiency 'metric' on the basis of policy context is that 'decision-makers' who are purported to have a clear goal, or set of goals, are likely to be conspicuous by their absence (Coast, 2004). A second major stumbling block in obtaining a suitable efficiency metric is the lack of universally agreed 'measure' of the diverse and often intangible effectiveness objectives of health care (Brazier et al., 2019; Mooney & Lange, 1993; Valtonen, 1993, p. 195; Wildavsky, 1969). A third issue in a triad of problems relating to efficiency indicators, which is of paramount importance, is the extent to which intangibles might impinge on the solutions to policy questions (Quade, 1967). These points will be elaborated upon in Section 6.

4.5 Producing information on 'efficiency' in practice

As noted in the previous sub-section, context, which often defines the goal towards which the efficiency indicator is aiming, is one important consideration when producing information on efficiency. However, it is important to keep in mind that there are numerous other important considerations, such as the credibility of the data or research underlying, i.e., the strength of the evidence on costs and effectiveness. When choosing the most appropriate way to operationalise the estimation of the cost per unit of effectiveness using CEA, there is necessarily a move from principles to practice, and also often from concepts and ideals towards constraints and pragmatism. The first main reason for such constraints and pragmatism is a

rather inconvenient truth for HEE; two major practical problems exist when undertaking health-economic evaluation: measuring the costs and the measuring the benefits (McPake et al., 2013, p. 80). Even if a goal is both known and universally supported, the best available methods of measurement and analysis may still thwart the assessment of that goal (Evans, 1984, p. 243). For example, practical limitations in the measurement of 'benefits' include difficulties surrounding the measurement of effectiveness, such as difficulties in observing effectiveness over time due to loss to follow-up, as well as a host of other difficulties in obtaining robust estimates of the relative effectiveness of different health technologies. A second reason for such constraints and pragmatism is that key health-economic evaluation texts and guidelines do not generally prescribe exactly what should be measured, i.e., what the numerator and denominator in ICERs should contain (see, e.g., (Neumann et al., 2016), (Drummond et al., 2015), (European Network for Health Technology Assessment (EUnetHTA), 2015), (National Institute for Health and Care Excellence, 2013), and (Ministry of Social Affairs and Health (Finland), 2003)). Despite valiant efforts it seems impossible to propose *uniform* context-independent guidelines or suggestions as to what might constitute the most appropriate measures of costs and effectiveness to use when applying CEA in health and social care (see, e.g., (Culyer, 2018), (Weatherly et al., 2017) and (Heintz et al., 2016)). Despite over 50 years of practicing health-economic evaluation, just as Steiner predicted (Steiner, 1965), there still seems not to be a single definitive methodology available. Therefore, the study presented here has approached the application of health-economic evaluation to the FinRSPC by adhering closely to the following advice:

> "So the first (and perhaps the most important) job of the analyst is to ensure that all feasible alternative methods of performing a task are under consideration from the outset, and that vision is not restricted by conscious or unconscious preconceptions about what is the best (or the only practical) way of doing things." (Williams, 1967)

It should be noted here that the use of such judgment and intuition in CEA can, in principle, be seen as quite normal for health-economic evaluation (see, e.g., (Quade, 1967), (Drew, 1967) and (Klarman, 1967)). Indeed, rather than being value-free or context-free evidence, or being a value-neutral or context-neutral process, economic evaluation is value-laden, i.e., normative, research (see, e.g., (Culyer, 2015) and Article III). The reason for this seems clear, each methodological guideline merely represents a collection of value judgments, even if these judgments are consensus statements by august, renowned or

erudite collections of experts (Mulkay et al., 1987). The review of the literature undertaken for this thesis would seem to strongly indicate that no practical economic evaluation can be definitive, each will be based on its own underlying perspective(s), each economic evaluation will be contingent on, e.g., the methodological value judgements, data, and assumptions used. Therefore, approaches to economic evaluation can indicate some of the implications of choices, but typically cannot provide a comprehensive evaluation (Dror, 1970). This being the case should not detract from the present study *per se*, but it does provide a strong rationale for taking these matters into account when interpreting the results of health-economic evaluation.

The current study will not follow the current dominant approaches of EWHEE or decision modelling, and reasons for this will be set out in the next section.

5 Applying 'economic evaluation' to the FinRSPC in practice

The general topic to which a form of HEE will be applied here is 'systematic mass prostate cancer screening based on prostate-specific antigen (PSA) testing', for brevity this will, hereafter, be referred to as either 'PSA screening', 'PSA-based mass screening', 'prostate-cancer screening' or 'screening for prostate cancer'. More specifically, CEA will be applied to a randomised controlled trial (RCT) of screening for prostate cancer, the Finnish Randomised Study of Screening for Prostate Cancer. This trial, and its continuous register-based follow-up, has been underway in Finland since 1996, hereafter this will be referred to as the 'FinRSPC'. In order to facilitate economic evaluation of the FinRSPC, postal questionnaires which included a range of questions both about health status and economic impact of screening were administered to men diagnosed with PCa in 1998, 1999, 2003 and 2011. Similar questionnaires were also sent to a trial subsample, consisting of 2,200 men: a random sample of 1,100 men from the screening arm and 1100 men from the control arm (see Article II (Booth et al., 2014a) for further details).

5.1 Collection and utilisation of materials

The start of the FinRSPC coincided with a period when the Finnish economy was boosted by a successful investment by Finnish government in mobile-phone technologies, namely in Nokia Ltd. and related businesses. This meant that time, resources and materials were, in principle, readily available for research. Prior to the turn of the millennium, obtaining data from registries was relatively easy for those working in the predecessor to THL, and data-protection processes controlling the use of questionnaire responses was demanding, but manageable, for researchers. Around the turn of the millennium there were changes in Finnish data-protection legislation and in the personnel conducting the health-economic research for the FinRSPC. After the first year or so of my involvement with the FinRSPC, a third round of

postal questionnaires was administered in 2004. A lack of Finnish language skills on my own part, as well as the aforementioned changes in rules governing the collation and use of registry and questionnaire data, contributed to an elongated application process for updating research permissions. In fact, the granting of permissions had to wait until a minor legislative reform in registry data handling had been implemented by the THL in 2010. During this period of waiting for permissions to link the questionnaire data and register data (2005-2010), preliminary analysis was undertaken on comparison of the information available from questionnaires and from those registers for which permission had already been granted. This analysis of data from the health-economic study alone showed fairly extensive opportunistic PSA-testing, often referred to as 'contamination', in the control arm. A large part of the review of the methodological literature for this study was also conducted in that period. After it was confirmed that all previous data could be used, a fourth round of postal questionnaires was undertaken in 2011. This fourth survey was undertaken whilst waiting for the administrative data itself on outpatient visits to be released from the participating hospitals (which finally arrived in 2012). The first article for the current study, mainly utilising the questionnaire data, was accepted for publication in 2012 and published in 2014. Between 2012 and 2015 a deeper understanding of the HTA literature was obtained during my work at THL, and various research permissions and register data were updated again between 2015 and 2017. In 2017 analysis of the registry data began in earnest and resulted in the publication of Article II in 2018, and the production of Article IV in 2018. The background research for Article III was developed between 2008 and its publication in 2019.

5.2 Literature review of economic evaluation of PSA screening

A literature search was conducted in April 2019 to find published studies related to the economic evaluation of PSA-based mass screening from the following databases: US National Library of Medicine PubMed database; databases created by the Centre for Reviews and Dissemination; Cochrane Database of Systematic Reviews; SCOPUS; Journals@Ovid Full Text; CINAHL and PsycInfo (via EBSCOhost); Web of Science and Tufts Medical Centre's CEA Registry (see (Booth, 2016) for further details of the search

strategy). The search was deliberately sensitive in scope, in an attempt to include all potentially relevant economic evaluation -related studies from these databases. On the other hand, e.g., after the search, exclusion criteria included systematic reviews of economic evaluations (Anderson, 2010). In addition, e.g., a number of earlier studies were excluded on the basis that the evidence from the two main randomised controlled trials on PSA screening was not yet available, both of which first reported their results in 2009 (Schroder et al., 2009) and (Andriole et al., 2009). Given the above, six relevant peer-reviewed articles describing economic evaluations were thus selected, from the 57 potentially relevant records identified through reviewing the searches from the above databases.

The following health-economic evaluations were selected from the search results (Shteynshlyuger & Andriole, 2011); (Pataky et al., 2014); (Shin et al., 2014); (Heijnsdijk et al., 2015); (Roth et al., 2016) and (Keller et al., 2017). These were all modelling studies, *none* of the studies located were similar to the current study, although some were based on data which, in part, came from earlier stages of the FinRSPC (Schroder et al., 2009).

It might already be clear from sub-section 4.5 that major differences can be expected between model-based health-economic evaluations or simulations using aggregate-level data and health-economic evaluations based on individual-level data from a single trial (Melnikow et al., 2013). For this reason, in an attempt to obtain further detail about the models and their results (in order to better assess them for credibility and comparability), e-mails were sent to the corresponding authors of the peer-reviewed publications identified. The corresponding authors of two of the articles identified by the systematic search responded and were willing to provide extra details (Pataky on behalf of (Pataky et al., 2014) and Heijnsdijk on behalf of (Heijnsdijk et al., 2015)), the other authors did not respond (with any useful information) to attempts to contact them. Although some of the modelling and the some of the data upon which the work by Pataky et al. and Heijnsdijk et al. were based, are open to scrutiny, in practice, as noted elsewhere, there seems to be a dearth of opportunities to fully assess models in many respects (Kaltenthaler et al., 2013). In part this may be the result of pressures to maintain a monopoly on the use of datasets within a research group or pressures to publish relatively short manuscripts in highly-regarded, peer-reviewed medical journals (Angell, 1986; Silverberg & Ray, 2018). However, it should also be noted that the development of tools or checklists for assessing the quality and congruence

health-economic-evaluation models is problematic, given the variation of approaches already noted in earlier sections (see, e.g., (Glick et al., 1994) or (Tappenden, 2014)). Due to the above, this thesis refrains from describing the *results* of the six studies located by this systematic search, given that they each fail, in one or more respects, to provide comparable information (Welte et al., 2004). Even with only the rather limited information about these six PSA-based mass screening studies, it is clear that the methods used in the located papers are testament to the pervasive heterogeneity in economic evaluations, especially when compared to the CEA presented as part of the current study. Therefore, any suggestion to compare the results would require extensive further investigation of the methods used in those studies, an investigation which, to a large extent, would likely be futile given their acontextual nature (Anderson, 2010). For example, all previous economic evaluations took clinical epidemiology as their cornerstone, in combination with assumptions, and it is that information which, to a large extent, drives their models. However, social epidemiology is likely to be of importance, too. When reviewing the empirical results from the current study it would be wise to note that "health is not just the product of health care, at the very least it involves environmental, economic and social factors." (Small & Mannion, 2005). For instance, screening for prostate cancer may appear to have had positive impact, in terms of prostate-cancer mortality, *on average*, but this does not mean that similar impacts of screening would be observed in all settings, or in all subgroups (see, e.g., (Cartwright & Hardie, 2012, p. 127)). Of course, ideally, analysis of the potential impact of socioeconomic status on outcomes would also be undertaken (see, e.g., (Ratcliffe & Gonzalez-Del-Valle, 1988), (Birch, 1997) and (Birch, 2002)). However, only in the last months of finalising analysis for the current study was more extensive socio-economic status data available from the statutory authority (Statistics Finland). Previously, socio-economic status data were only available for diagnosed men, thus preventing potentially crucial analysis of the potential impacts of socio-economic status.

Failure to take considerations such as those noted above into account may lead to overly parsimonious models, which fail to focus on those types of uncertainty which are more commonly acknowledged to be harder to assess (Briggs et al., 2012). Treating the underlying clinical epidemiology as a given also tends to exclude the possibility that they could provide novel contributions to underlying epidemiological debates (Glick et al., 1994), or be able to take into account important environmental or socio-economic factors

(Kothari & Birch, 1998). Of course, there is a recognised tension between extending the features or structure of models and obtaining the data or estimates needed to populate a model in a more meaningful manner (Roberts et al., 2012), as can be seen from the following quote:

> "Calculating quantitative solutions using the wrong criteria is equivalent to answering the wrong questions. Unless operations research develops methods of evaluating criteria and choosing good ones, its quantitative methods may prove worse than useless to its clients in its new applications in government and industry." (Hitch, 1953)

According to Glick et al. (1994), it is often stated, apparently often without justification, that information based on synthesis and modelling offers improvement over information from any single RCT. The claim is made here that, rather than relying on dogmatic reasoning about the usefulness of modelling, a more nuanced approach should be taken. This would include broader acknowledgment that one goal of analysis is to find optimum solutions, whilst *assuming* that appropriate predictions can be made (Lessard & Birch, 2010). When confronted with data or assumptions that contradict the generalisability or transferability of modelling, e.g., as set out in the current study, there would seem to be a moral duty for proponents of modelling to provide empirical support for such assumptions. The view taken here is that the results of the CEA of the FinRSPC cannot easily be compared to results from previous modelling studies, because although they represent the same general construct, they do so in fundamentally different ways. Given the difficulties, in general, in the evaluation of the effectiveness of prostate-cancer screening, a more thorough investigation and assessment of the usefulness of modelling should be undertaken. However, that was not possible within the scope of this thesis. Of course, methodological choices are of importance, such choices may be made on the basis of, amongst other things, the data available, ideological proclivities or pragmatism (see, e.g., (Naylor, 1996), (Krahn, 1996) and (Russell & Sinha, 2016)). The further claim is made here that the context specific nature of the FinRSPC results should not be forgotten, there are numerous reasons why it might not necessarily be prudent to generalise from the findings of either the ERSPC or the FinRSPC (Pawson, 2006, p. 180). Indeed, it is noted here that the findings in Article IV about the impact of screening for prostate cancer in terms of prostate-cancer mortality are of similar direction, but of smaller magnitude, than those found elsewhere in the ERSPC (Hugosson et al., 2019). Despite the fact that Article IV does not incorporate an estimate the HRQoL impact of screening for prostate cancer,

research seems to indicate that there is a strong interrelation between any prostate-cancer-mortality benefits associated with screening for prostate cancer on the one hand, and the unnecessary diagnosis and harms of treatment on the other (Auvinen et al., 2016). A rather cursory reading of the modelling studies suggests they do not undertake extensive sensitivity analysis, even given wide-ranging evidence on the possible extent of length bias, lead-time bias or overdiagnosis. For example, in one of the six studies identified by the review (Heijnsdijk et al., 2015), it appears that extensive testing of the sensitivity of results to less conservative assumptions about the potential impact of screening on overdiagnosis was not undertaken. Further, although the modelling studies located by the systematic search typically do attempt to validate the results of modelling exercises against observed prostate-cancer mortality, they do not appear to attempt to validate their results with respect to either observed HRQoL outcomes or costs.

Another potentially important factor which modelling studies could investigate further is any PSA testing which takes place outside systematic screening programme, i.e., opportunistic PSA testing. Limited sensitivity analysis on this factor seems odd, even given a relative dearth of evidence on opportunistic testing from epidemiological studies between publications by Ciatto et al. (Ciatto et al., 2003) and Kilpeläinen et al. (Kilpeläinen et al., 2017). After reviewing the retrieved literature it seems that many model-based analyses utilise a comparator of 'no screening' rather than the *de facto* situation in many jurisdictions of unorganised or non-systematic screening, often referred to as opportunistic PSA testing. Despite controversy over the impact of opportunistic PSA testing on the effectiveness estimates coming from studies of prostate cancer screening (see, e.g., (Arnsrud Godtman et al., 2015)), opportunistic PSA testing in the control arm seems likely to dilute the relative effects reported by PSA-screening trials (Roobol, 2015). One other pitfall in the use of modelling is that the focus of intellectual effort is on the model, rather than on the policy question (Quade, 1967). For example, there is a tendency to focus on describing the effect of stochastic uncertainty via sensitivity analysis within most models, rather than on structural uncertainty, which is less amenable to description (Afzali & Karnon, 2015). Modellers may also choose to accept, perhaps somewhat unthinkingly, the information stemming from 'disciplines' such as clinical epidemiology. This brings with it the potential of drawing attention away from any potentially important effects of screening which are not addressed by clinico-epidemiological research.

In summary, taken together with findings from the current study, the observations above cast some doubt on the credibility of the six studies located by the systematic search. Further, the ability to sensibly compare the approaches of the identified studies, to the approach taken by the current study, is also limited. For these two reasons, no further extensive comparison between the results of these dissimilar analytical strategies will be entertained here.

5.3 Pragmatic considerations in the CEA of the FinRSPC

While waiting (between 2005 and 2010) for the relevant research permissions to be granted, and while waiting for delivery of the bulk of the cost-related register data in 2011 (from THL and Kela), time was available to ponder the many different potential approaches to health-economic evaluation which could be used with the data available from the FinRSPC. Of course modelling, either with the HRQoL data from the FinRSPC questionnaires or with HRQoL evidence from the literature would be possible, but given the nature of the topic this would involve the use of strong assumptions and would not necessarily make best use of the available register data (Sheldon, 1996).

On balance, given that a substantial period of real-world follow up had been amassed, the decision was made to undertake a wholly register-based CEA in order to highlight the 'real-world' evidence on the costs, effects and cost-effectiveness of the FinRSPC. This application of CEA to the FinRSPC will be described in the section after next. Before that, however, it will be useful to briefly present the findings from the review of the methodological literature with respect to the conduct of health-economic evaluation.

5.4 CEA as part of iterative application of economic evaluation

As seen in the previous sub-section, research relies on choices made by the analyst(s) about data to be used, the approach to economic evaluation and associated assumptions. A policy question must first be translated into a research question. Of course a direct link can be drawn between the chosen

research question (often referred to as a 'decision-problem', when used in a decision-analytic sense) and the *policy* question (the latter often framed as being holistic in nature, i.e., akin to holistic evaluation). However, it is widely recognised that a link between the two should only be made at the discretion of a decision maker, if at all. One aside to note here is that when economists conflate a research question and a policy question, they run a grave risk of "repudiating their claim to any independent norms of economic expertise" (Mishan, 1982b, p. 43). To try to help to clarify the above concerns Figure 1, below, summarises my reading of the existing health-economic evaluation literature and suggests that evaluation processes should be iterative, and the answers to research questions should not automatically be considered as single-shot solutions to policy questions. It should be noted from the figure that there is not necessarily a direct link between the 'decision problem', i.e., the research, on the one hand, and the policy question on the other. Given abstractions typically made in research, the separation of the policy question and the iterative cycle of evidence generation in Figure 1 is intentional: it is justified by, e.g., the observation from the literature that it is generally not sensible to conflate health-economic evaluation with the solution to any policy question (see, e.g., (Williams, 1991)).

Any 'economic' research question, on the topic of screening for prostate cancer at least, would tend to be an abstracted, partial or truncated one, thus ensuring separation from the policy question *per se*. All new information added to the evidence base as part of the cyclical process of evaluation would, *per se*, seem to require assessment and any appropriate qualifications. The idea here is that the cycle of evidence generation would ideally gravitate towards the policy question over time, but this may take multiple iterations and may be thwarted if the focus of some research is less than wholly inappropriate, or if the research questions are unanswerable due to lack of appropriate data or extant research evidence. As with any attempt to provide robust information, or improved approaches to policy-planning processes (Dewey, 1929, p. 41), the use of health-economic evaluation should be promoted by a suitably resourced multidisciplinary team.

Figure 1. A proposal for at least six stages in an iterative cycle in the contribution of economic evaluation in informing holistic evaluation (mainly developed using (Tugwell et al., 1986), *Figure 4.1*, as well as (Quade, 1971a), *Figure 2*, (Sculpher et al., 2006), *Figure 1* and (HM Treasury, 2018), *Figure 2*.

In line with the above proposal, as well as with earlier suggestions (see, e.g., (Cochrane, 1972 (2004))), the economic evaluation of the FinRSPC is an attempt to provide additional information to the existing body of evidence on screening for prostate cancer. It is not the aim of the current study to undertake a wide-ranging HTA (as set out, e.g., by (The European network for Health Technology Assessment (EUnetHTA) project, 2016)), but hopefully the current study can provide some useful evidence for use in research in future.

5.5 Choosing the type of economic evaluation for the FinRSPC

According to the original trial protocol HRQoL, costs and CEA are the secondary foci of the FinRSPC (Miller et al., 2001). The first piece of health-economic information, Article I concerning health-economic HRQoL in the FinRSPC, was published after the last set of health-economic questionnaires was collected in 2011 (Booth et al., 2014a). Only one article appears to have been published about HRQoL in the FinRSPC after that (Vasarainen et al., 2013), although at least two more are currently being processed for peer-review. An article relating to the second health-economic component, i.e., costs associated with the two arms of the FinRSPC, was published after the registry data became available in 2017 (Article II). Given the above timelines for the first two components of this health-economic study, in combination with the drafting of Article III, the process of deciding what type of analyses would be undertaken was a gradual one. Sub-sections 5.3 to 5.5 of this thesis provide a rationale for the choice of the type of health-economic evaluation undertaken here, and some further reasons for the eventual choice of methodology can also be found in Article III. The decision was made with useful hindsight about the extent and qualities of the data available from the FinRSPC, and with awareness of the subjective nature of judgments concerning how to approach health-economic evaluation (Quade, 1967). For instance, after much deliberation, the deficiencies of the questionnaire data suggested that little robust evidence on HRQoL would be available from the FinRSPC itself. However, although it is always possible to try and incorporate a wide range of data from various sources, at some point potentially relevant data for future models also has to be *produced*. Given the efforts of the FinRSPC research group, the efforts necessary to obtain the register data, as well as the completeness of coverage (but not of scope) of the register data, the multidisciplinary decision was taken to concentrate analysis on the most robust source of information, however incomplete in terms of scope.

As the main published evidence concerning the primary outcome of the FinRSPC as a stand-alone trial (mortality from prostate cancer) was over five years old (Kilpeläinen et al., 2013), the health-economic evaluation of the FinRSPC updated those results. No other articles have so far been published about the cost-effectiveness of the FinRSPC. After describing the FinRSPC in the next sub-section, the subsequent four sub-sections will briefly summarise and review the four FinRSPC-related articles.

5.6 Background information related to the FinRSPC

This sub-section briefly describes the nature of the FinRSPC study, its design, methods and results related to the health-economic evaluation of the FinRSPC. In order to do that, it may be useful for the reader to understand, in simple terms, a little more about screening for prostate cancer (PCa). In the current study the term 'screening' is used to refer to PSA testing of the kind that is undertaken as part of an organised, systematic, public-health screening programme, often referred to as 'mass screening' (Hakama, 1991). PSA screening can be classed as an early-detection screening test, which aims at diagnosing early-stage cancers, in principle to subject them to life-prolonging treatments or even potential cure. Diagnosis of PCa has increased rapidly in the past 25 years or so in Finland (see Figure 2), just as in most industrialized countries (Auvinen & Hakama, 2017).

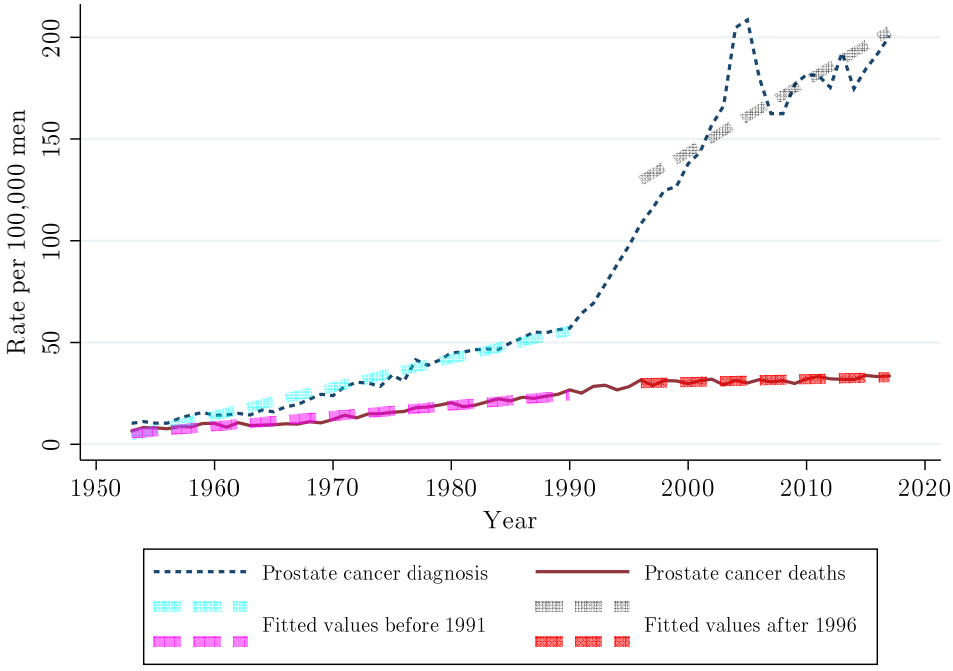

Figure 2. A figure showing the development of prostate-cancer incidence and mortality in Finland (Source data: https://cancerregistry.fi/statistics/cancer-statistics/)

Given that the usual rationale for investigating screening for prostate cancer, is that it could detect early cases of invasive (or pre-invasive)

prostate cancer, with the aim being to decrease *prostate-cancer* mortality (Auvinen & Hugosson, 2003), ideally, this rationale should be thoroughly tested (Malm, 1999), at the very least due to the potential harms associated with overdiagnosis. Overdiagnosis is an important consideration in all forms of screening, and, e.g., its definition can include the anxiety associated with being labelled with diagnosis of cancer, any harms which result from being subjected to intensive surveillance for any advancement of the disease (Kalager et al., 2018), as well as any lack of benefit, unnecessary cost or harm from treatment of an overdiagnosed case (Loeb et al., 2014). More will be said about overdiagnosis below.

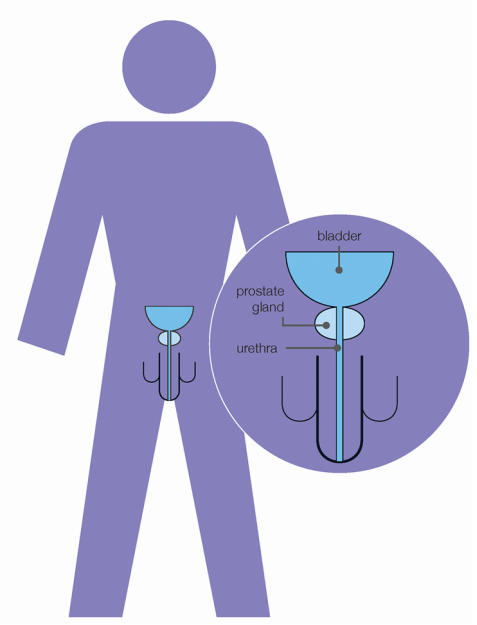

Figure 3. A graphic depicting the location of the prostate (courtesy of Prostate Cancer UK)

The prostate is a gland and on average, in healthy adult males, is typically similar in size to a walnut. It is located under the bladder in males and wraps round the urethra (see Figure 3). Its size tends to increase with age and, although it can also become enlarged due to benign and/or malignant disease, there is much variation between individuals. The prostate itself is more susceptible than many other organs to development of cancer, however it is rare to die of PCa before the age of 55 (currently less than 1 in 100,000 Finnish men per year). Most men over 75 will have some cancerous lesions in their

prostate and most of these men would be unaware of its presence, given PCa is often asymptomatic in its less virulent forms. Therefore, unless men seek out PCa using a medical technology such as PSA testing or magnetic resonance imaging (MRI) (Bell et al., 2015), indolent forms of PCa typically remain undetected. PCa is thus relatively common in older men and, although it poses no serious threat to most men, for some, PCa will be life-threatening or fatal.

Prostate-specific antigen (PSA) is a protein produced exclusively by the prostate, and PSA in the bloodstream may indicate disease in the prostate, be it cancerous or benign. By the end of the 1980s the testing of serum PSA levels was increasingly being used to identify otherwise undetectable prostate cancer (Stamey et al., 1987). PSA not a tumour marker, i.e., there is no level of PSA that would definitively indicate the presence or absence of PCa (Haythorn & Ablin, 2011). On the one hand, if serum PSA is low ('low' PSA is typically defined in relation to the size of the prostate) it means that aggressive prostate cancer is less likely, even if it does not rule it out completely. On the other hand, high PSA can provide an early warning for PCa and the chance to pick cancerous lesions up early, particularly for men who have a more aggressive and significant form of the disease (Manvar et al., 2013).

The potential benefit of a screening programme is related to the lead-time 'gained' by early diagnosis, but this relationship is not straightforward, e.g., because of the interrelation between lead-time and overdiagnosis (Finne et al., 2010). To understand cancer overdiagnosis, it is first useful to understand the heterogeneity of cancer progression, which is first highlighted by a quote, and then depicted in Figure 4, below (modified from Figure 1, Welch & Black (2010), reproduced with permission).

> "Cancer is not a single entity. It is a broad spectrum of diseases related each other only in name." Crile (1955)

In line with the above quote, one of many definitions of overdiagnosis relates directly to the identification of one end of a spectrum of prostate cancer, i.e., those slow-growing or non-progressive cancers which would only rarely spread or cause harm, even though 'correctly' diagnosed (Carter & Barratt, 2017). Screening can easily lead to overtreatment and harm, in addition to overdiagnosis, in the case of such very slow-growing or non-progressive cancers. Of course, screening, when followed by successful treatment for more rapidly-growing cancers, can prevent death from prostate cancer, albeit with the potential for accompanying detriments to health-related quality of life. On the other hand, some rapidly-progressing cancers may not

benefit from treatment, or treatment may have minimal beneficial effect, with death from prostate cancer occurring even with early treatment (Esserman et al., 2014).

Simply put, one 'opportunity cost' of PSA screening seems immediately apparent: although in the *absence* of PSA screening, more men would likely die from prostate cancer, in the *presence* of PSA screening, more men would likely be overdiagnosed and overtreated. In addition, there may well be accompanying negative effects on the health or well-being of the men for whom death from prostate cancer is avoided, in addition to any positive effects.

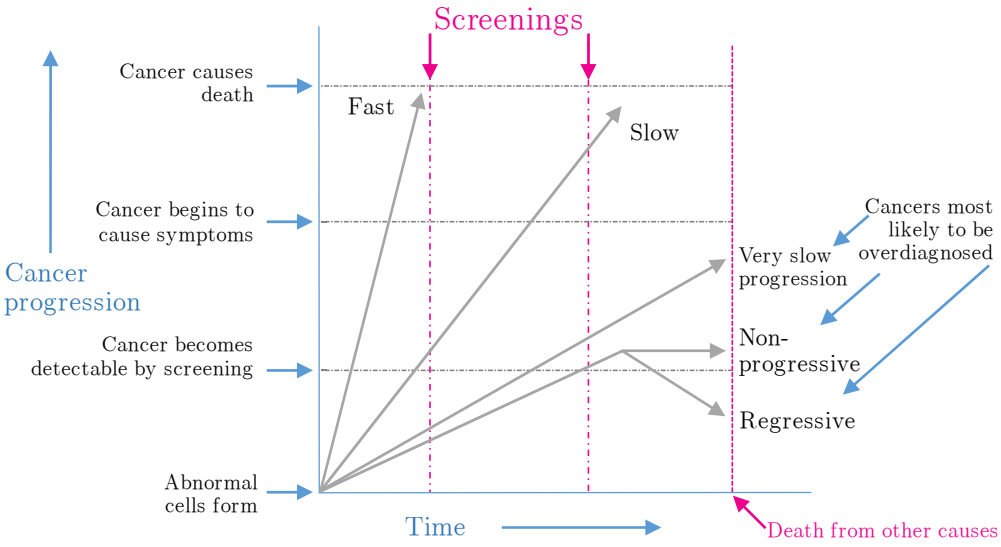

Figure 4. A graphic depicting the heterogeneity of cancer progression

Because the rate of progression of prostate cancer can vary so widely, this creates a major problem in quantifying the harms or benefits of screening, at least in terms of the presence or absence of overdiagnosis, which, in turn, may also severely restrict comprehensive evaluation of the harms and benefits of treatment in trials of prostate cancer screening. In fact, although overdiagnosis directly affects individuals and their families, the quantification of overdiagnosis requires drawing inferences at a population level, modelling or other forms of estimation (Carter et al., 2016). It should be noted that a wide range estimates of prostate-cancer overdiagnosis have been made, which range from 1.7% to 94% (see, e.g., (Loeb et al., 2014) and (Welch & Black, 2010)). Because of a lack of robust information about the extent of

overdiagnosis, e.g., in the FinRSPC, the ability to assess even lower-level 'opportunity costs' of PSA screening in a straightforward manner is restricted.

Further, it should also be noted that, during the period when men were still being invited to participate in FinRSPC screenings (1996-2008) three developments occurred. Firstly, there were changes in the application of pathologic grading of tumour differentiation. Secondly, there were changes in the surgical methods used, and indications, for radical prostatectomy. Thirdly, contamination, in the form of opportunistic PSA-based testing, increased year on year.

The first of these may have impacted on men through potential changes in over-detection, the second may have impacted on men through, e.g., potential changes in over-treatment. To reiterate, some of the main unwanted consequences of treatment for PCa have been related to radical prostatectomy, including loss of sexual function or incontinence. In a related development, active surveillance has been widely adopted and aims to individualise the treatment of low-risk prostate cancers (Hardie et al., 2005). Active surveillance is likely to have decreased the overtreatment of indolent tumours by aiming to provide treatment only for those men with clinically significant cancers, although its impact currently seems somewhat limited according to one study (Drost et al., 2018). The third of the changes over time was the increasingly large proportion of men in the control arm who had had a PSA test (Kilpeläinen et al., 2017). Reasons for the increased contamination include the prevalence of age-cohort health checks as part of fairly wide-reaching occupational health care packages in Finland, as well as, e.g., direct-to-consumer marketing of PSA testing in shopping centres.

The study population of the FinRSPC consisted of men aged between 55 and 71 and resident in the study areas. On the first day of each year from 1996 to 1999, 8,000 men were assigned at random to the intervention group (the screening arm to which systematic invitations to PSA-based mass screening were planned) (Finne et al., 2003). The remainder of the age cohort each year were assigned to the (larger) reference group (the control arm to which no invitations to PSA-based mass screening were sent). As the end result of extensive organisation and planning, the FinRSPC intervention offered systematic invitations to PSA-based mass screening for men in the screening arm, and a standardised, pre-planned approach to clinical follow-up and potential diagnosis for screen-positive men. Men were invited to PSA screening a maximum of two or three times, at four-year intervals (i.e., if men in the

screening arm were aged over 71, or if they had otherwise received a PCa diagnosis, they were no longer sent an invitation to participate). The first screening round was carried out between 1996 and 1999, the second between 2000 and 2003 and the third between 2004 and 2007. Further details about the study design, methods and results of the FinRSPC are available from Article IV, as well as from (Kilpeläinen et al., 2013), Article I and Article II.

As already noted in Sub-section 5.4, the research presented here is not a HTA, nor does it represent a holistic evaluation of a mass-screening programme. Indeed, it should be noted that, contrary some of the principles found in the widely-referenced guidelines on screening evaluation (such as (Wilson & Jungner, 1968), (Butterfield, 1968), (Pole, 1968), (Cochrane & Holland, 1971) and (The Danish Council of Ethics, 2001)), the aetiology and natural history of prostate cancer are still not well understood. There are only weakly-identifiable associations between, e.g., diagnosis of prostate cancer and socio-economic status; use of certain pharmaceuticals; hereditary lineage; molecular or genetic makeup; or environmental factors. In light of this dearth of current knowledge, the current study attempts to focusses on the health-economic information which may still add *some* useful information to the evidence base, in the hope that the findings herein may be better investigated in future, or lead to better future research and better-informed decisions.

5.7 'HRQoL': review of study design, methods and results

As there was a dearth of evidence of the potential long-term impact of PSA screening for prostate cancer on health-related quality of life (HRQoL), the objective of Article I was to report the findings gathered from some of the health-economic data collected as part of the FinRSPC on HRQoL. This attempted to add to the evidence base using postal questionnaire surveys which were conducted in 1998, 2000, 2004, and 2011 among men from a random subsample of the trial population ($n = 2200$), as well as among men in the FinRSPC diagnosed with PCa (eventual total $n = 7011$, which included all available men prior to each survey). In 2011, for example, 1587 responses were received from men with PCa in the screening arm and 1706 from men with PCa in the control arm. In addition, e.g., in 2011, from the trial

subsample, 549 men in the screening arm and 539 in the control arm provided responses.

Before looking at any analysis from this CEA of the FinRSPC, it is important to highlight a few potential causes of bias in any analysis which does not take its 'time origin' as the point of randomisation (Duffy et al., 2008). Thus, even *within* an RCT, length bias and lead-time bias may affect any analysis in which the starting point for analysis, or time origin, is taken as the point of diagnosis (Baker et al., 2002). Although randomisation occurred before consent in the FinRSPC's population-based age cohort, thus likely protecting analyses from biases due to selection or volunteering, important biases related to screening, such as length bias and lead-time bias, are likely to cause problems for interpretation when the analysis did not use the date(s) of randomisation as the time origin(s).

Analysis, from the point of randomisation, comparing health-state-value scores in the intervention and control arms using three HRQoL measures (15D, EQ-5D, and SF-6D), was undertaken on questionnaires administered in 2011 to the random sub-sample of men from the trial (Booth et al., 2014a). In this study, Article III, the average scores calculated from all three HRQoL measures were higher in the control arm, but only the EQ-5D measure indicated small, statistically-significant advantage in favour of the *control* arm (control vs. screening arm: EQ-5D: 0.85 vs. 0.83, $p = 0.039$). This finding would seem to suggest PSA screening might have reduced the health-related quality of life of men in the screening arm in genereal, at the last available point of HRQoL follow-up in this study. One other comparison of HRQoL, undertaken between patients in a randomised study of treatment and a population-based control group of men matched for region and age (Johansson et al., 2011), found that significantly more men reported moderate to high levels of anxiety as patients, than did men in the control group. Although the study by Johansson et al. did not specifically investigate screening, it did focus on corollaries of screening, in terms of the effect of treatments on *some* important indicators of HRQoL. In addition, if diagnosis leads to treatment with curative intent, e.g., in the form of radical prostatectomy or radiotherapy, there is a risk of lifelong treatment-induced diseases and states, such as problems relating to urinary or anal leakage, in addition to commonly-occuring erectile dysfunction (Steineck et al., 2019).

Ideally, it would seem reasonable to suggested that future evaluations of PSA screening would attempt to concentrate on identifying differences for

those men with the biggest (potential) impact (Auvinen et al., 2017), i.e., men diagnosed with prostate cancer. However, it may well be the case that there is no available method which can be used to correct for length- or lead-time biases *within* an RCT of prostate-cancer screening, i.e., when attempting to analyze outcomes for diagnosed men. For instance, methods do not appear to be available which can accurately adjust for there being a longer period of time when cancers could be found via screening (length bias) or for the follow-up time from screen detection being longer than the follow-up time from clinical detection (lead-time bias) when analyzing outcomes for diagnosed men. The above problems would seem to thwart not only the comprehensive evaluation of HRQoL effects related to prostate-cancer screening, but also the comprehensive evaluation of costs and cost-effectiveness, as well as comprehensive evaluation of prostate-cancer screening *per se*.

With such hindsight, it seems impossible for definitive conclusions to be drawn from *any* analysis presented in Article I, other than for those results relating to the random sub-sample of men from the trial. This is not only because the patient data were collected without regard for the point in time when each man had been diagnosed; but also, and perhaps more importantly, results pertaining to diagnosed men were reported without taking into account the likely problems with these results due to length bias or lead-time biases. Information in order to fully comprehend the risks of overdiagnosis is still lacking (Dempsey, 2019), because the currently available evidence may be unable to adequately inform men or organisations facing choices over prostate-cancer screening, of all significant benefits or harms.

Further to the above, a brief exchange occurred in the literature relating to Article I (Bergman & Litwin, 2014; Booth et al., 2014b). Subsequent to those discussions it seemed clear that generic HRQoL measures commonly used in health-economic research and postal administration of questionnaires may not be well suited to long-term follow up of men. Reasons for this include the potential influence of factors unrelated to the prostate-cancer screening trial such as aging; impacts on health unrelated to the trial; socio-economic status; and health-related non-response to questionnaires; all of which have the potential to skew the results of any HRQoL research, including all analyses included in Article I.

5.8 'Costs': review of study design, methods and results

Little real-world evidence exists on the impact of prostate-cancer screening on healthcare costs, despite cost-related information often being considered as a necessary prerequisite to fully-informed evidence-based mass screening policy (The Danish Council of Ethics, 2001). Article II estimated differences in register-based average costs of publicly-provided healthcare in each arm of the FinRSPC at 20 years.

Article II used individual-level register data on prescription medications, as well as on inpatient and outpatient care, in order to estimate healthcare costs for the 80,149 men during the first 20 years of the trial and compared average healthcare costs for a number of sub-groups by trial arm. Despite this long-term follow-up of the large FinRSPC age cohort, the 20-year period may be too short, or the inherent heterogeneity in costs at the level of individual patients too great, to provide robust information about the financial burden of prostate-cancer screening to parts of the Finnish health-care system. The conventions of Finnish, article-based doctoral theses largely preclude the introduction of new results related to the constituent articles. Therefore, it must suffice to note that the estimates reported in Article II are problematic because most of the analyses do not take account, amongst other things, the potential impact of lead time bias, length bias or possible variation in treatment practices.

In addition, journals tend to favour the reporting of between-trail-arm differences in average costs at the expense of other results, even when this neglects the fact that some of the relevant formal analytic information is the entire probability density, rather than only the mean (Manski, 2019). Due, for example, to academic journals having word limits which may restrict what is allowed to be reported, there often remains relatively little space in many peer-reviewed medical journals for sensitivity analyses. Simplifying any message can easily do an injustice to hard-earned data or time- and resource-consuming analysis. Longer-term follow-up and wider-ranging research would be required to be better informed about what factors might drive the real costs of, or savings from, the potential introduction of PSA-based mass screening.

5.9 'CEA': review of study design, methods and results

The register-based 'CEA' of a 20-year-long pragmatic randomised trial of PSA-based mass screening provides new evidence on costs, comparative effectiveness and cost-effectiveness. At the time of writing this thesis, Article IV appears to be the only economic evaluation of PSA-screening which combines the power of a randomised, controlled trial with extensive follow-up via real-world data from comprehensive statutory health-care registers (Booth et al., 2019). In line with previous studies, the reported findings do not lend support to the idea that PSA-based mass screening in the FinRSPC could be classed as 'cost-effective' overall. Our study finds that health-care costs seem quite similar between the arms on average, even when account is taken of the screening-related costs *per se*. However, again on average, just over one man per 1,000 in the screening arm avoided death from prostate cancer during the trial (to the end of 2015), and also that, on average, around two additional men, per 1,000 men in the screening arm, died. An infographic is included here, which is intended to convey, in a concise manner, the main screening-related outcomes of the FinRSPC after 20 years of the trial (Figure 5, p. 63). Although it should be stressed that no single piece of information can necessarily represent the complicated or complex nature of the results of any evaluation of prostate-cancer screening, some interpretations can still tentatively be drawn from either the infographic, or the ICERs reported in Article IV.

One interpretation is that systematic prostate-cancer screening is associated with a small reduction of prostate-cancer mortality from 1.1% to 1.0%. In other words, on average, one man per thousand men systematically screened has been spared death from prostate cancer during the 20-year follow-up period of this trial. Despite there being a *potential* for misclassification of causes of death for men within the trial, it should be noted that research has suggested that the accuracy of classification was high in the FinRSPC (at least prior to 2003) (Mäkinen et al., 2008), thus it seems unlikely that the observed differences would be the result of 'sticky diagnoses' (Black et al., 2002). However, this interpretation should not be taken in isolation. In addition, the fact that overall mortality is found to be higher in the screening arm (although not statistically-significantly so) *could* be a cause for concern; because it would be expected, *a priori*, that if screening reduces prostate-cancer mortality it would also reduce (at least ever-so-slightly) overall mortality.

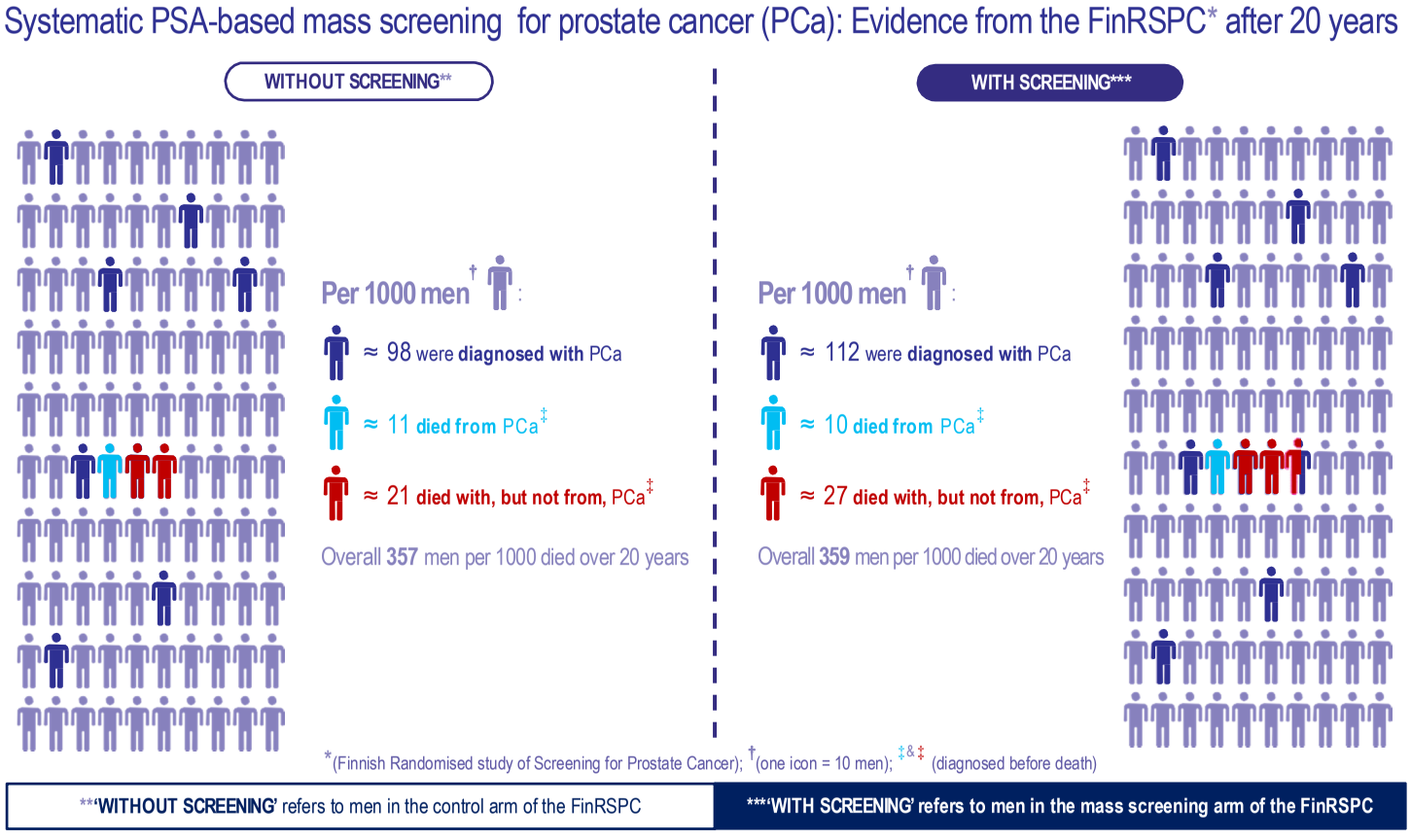

Figure 5. Infographic depicting some of the findings from the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC), after 20 years

Of course the observed differences could simply be due to random variation in mortality, especially given that the two arms of the FinRSPC contain many men. There have only been two other population-based trials of prostate-cancer screening where randomisation occurred before consent, one in Italy and one in Gothenburg (Hugosson et al., 2010). Only the Gothenburg trial has published separate, peer-reviewed results for all-cause mortality, and although in that trial (after 10 years) over three prostate-cancer deaths were prevented per 1000 men screened, overall mortality was higher in the screening arm (albeit by the smallest possible margin, with a 10-year mortality in the control arm of 1,982 versus 1,981 in the screening arm). It is useful to note here that although comparisons between all men in each arm would be ideal, in order to robustly analyse all-cause mortality our study could need to be about five times bigger (Heijnsdijk et al., 2019), and this would not have been practicable in Finland in the 1990s, because the men in the FinRSPC already made up 40% of Finnish males of that age at that time.

Of course, as noted in sub-sections 5.7 and 5.8, one interpretation of the results presented in Article IV is that the reported results relating to diagnosed men are much influenced by length bias or lead-time bias, and as such the reported results may not offer clear guidance, or *only* robust findings, to any decision-maker. But it might also be useful to note that cohort studies have suggested the possibility of an increased risk of death from suicide and cardiovascular death 30 to 90 days after prostate-cancer diagnosis (see, e.g., (Fall et al., 2009) and (Fang et al., 2010)).

Although the secondary analyses reported in Article IV should not be regarded as robust, they may still be useful as they can likely give one indication, or estimate, of the extent of overdiagnosis in the trial. One line of reasoning for this runs as follows: In the FinRSPC, after 20 years of follow-up, around 14% *more* men in the screening arm were diagnosed (a 1.14-fold cumulative incidence: with 11.9% of men diagnosed in the screening arm compared to 10.5% in the control arm, after 20 years of follow-up). In addition, around 1‰ of men are spared death from prostate cancer in the screening arm (around 1.0% in the screening arm compared to around 0.9% in the control arm, after 20 years of follow-up, see Fig 1 in Article IV). Hence, given that length of follow-up, around 14 more men needed to be systematically screened and diagnosed to prevent one prostate cancer death. As, on average, with the follow-up available to the current study, for every 14 men diagnosed, around one death from prostate cancer was averted, ten or so were still alive, and

around three had died of other causes. If the trends in mortality during the last 20 years would continue, then for every diagnosed man who is saved from prostate-cancer death, there will likely be two or three diagnosed men that die of other causes, remembering that cause-of-death statistics suggest that PCa mortality is around 3% (see Figure 2). If prostate-cancer screening in the FinRSPC would continue to prevent around 13% of PCa mortality, in line with the FinRSPC 20-year results, it might reduce PCa death overall from around 3% to around 2.7% (by the time all men in the FinRSPC cohort have died). In that case, although three less men per thousand might die of prostate cancer, at the same time 11% or so, instead of 10% or so, of men from the FinRSPC cohort would have likely been diagnosed as a result of systematic screening (i.e., around 14 *more* men per thousand might be diagnosed). According to the above relatively crude reasoning, one estimate of 'overdiagnosis' would be the odds of being diagnosed with, but not dying of, prostate cancer, i.e., around 11:3 or 10:4, a probability of around 80% or 70%, respectively. Although these crude calculations lack precision, in part due to the waning effect on mortality attributable to the FinRSPC's intervention, such estimates would seem to provide one potential *indicator* of overdiagnosis, in terms of one of the most robust forms of outcome measure, i.e., mortality.

Given the current lack of published research related to better understanding the effects and extent of overdiagnosis and other biases within trials of prostate-cancer screening, extending epidemiological studies to better address this dearth of information may be able to usefully supplement the existing evidence base. It is claimed here that if the assessment of *all-cause* mortality is thwarted due to lack of statistical power and if epidemiological analysis is focussed primarily on prostate-cancer mortality, without due regard for problems posed to evaluation by other outcome measures, related to overdiagnosis, as well as other biases, then subsequent economic evaluation may have weak foundations. Indeed, any weaknesses in the evidence base with respect to length bias, lead-time or overdiagnosis bias would also likely affect estimation of the impact of prostate-cancer screening on HRQoL and costs, as noted in earlier sub-sections, again potentially constraining the ability of simplistic CEA models to reliably *inform* decision-making.

Given the above, solid evidence directly from randomised controlled trials of prostate cancer screening seems likely to remain sparse. Extending both the scope of analysis and the follow-up period *may* help to produce more definitive or informative results, but, e.g., gathering of further information on HRQoL

of men in PSA trials seems likely to remain deficient. Suggestions in recent clinical guidelines that men should be informed of, and understand, the benefits and risks of screening would seem difficult to justify, given a reliance on current, potentially *relatively uninformative*, estimates of benefits and risks (US Preventive Services Task Force, 2018). For example, the 20-year finding that the FinRSPC appears to be associated with high levels of overdiagnosis and with increased overall mortality could be interpreted as implying that decisions about prostate-cancer screening made prior to the appearance of such information were relatively poorly informed. One of the potential contributions of evidence-based medicine seems absent; because previously-published studies do not typically clearly inform stakeholders about the lack of evidence. Without appropriate qualifications relating to the evidence base, it is more difficult to be aware of the uncertainties about decisions (Tunis, 2007). The 'best available evidence' may be unable to make men or organisations facing choices over prostate-cancer screening adequately aware of all possible benefits or harms, indeed, information in order to fully comprehend the risks of overdiagnosis may always remain inadequate (Dempsey, 2019).

The way the FinRSPC is analysed here is as a pragmatic, descriptive or exploratory study, not as an explanatory study. Further analysis and more data are both beyond the scope of the current study, but either or both would likely be necessary to understand how and why the screening intervention worked or failed to work (Clarke, 2006). The results of the FinRSPC may suggest that prostate-cancer screening has had a variety of impacts, but there is no way of deducing what it is that produced these results without costly additional analysis (Pawson et al., 2005). When the main effects of an intervention are not known with certainty (as in the case of prostate-cancer screening), but merely presumed, then a relatively small exploratory RCT may not be best equipped to contest those presumptions. The current study, which has novel findings from a RCT, could offer important information about prostate-cancer screening to supplement that from earlier studies. However, given these are the first findings in this direction from a RCT, the information should not be considered robust in terms of scientific reproducibility, but it may usefully add to earlier evidence from cohort studies (see, e.g., (Fall et al., 2009)). However, we should not draw conclusions too soon, corroboration or contradiction from similar analyses from similar mass-screening trials should be sought first. On the other hand, in the presence of ambiguity about such a

potentially fundamental part of the overall impact of screening, warnings about the potential for harm may well be appropriate. Indeed, another fundamental problem should also be kept in mind when evaluating cancer-screening trials, given that they can include a high amount of 'noise' relative to the 'signal' of the screening effect, potentially requiring studies to include huge numbers of men in order to facilitate comprehensive evaluation (Baker et al., 2002).

In addition, it should be noted that when interpreting any results from Articles I, II and IV, the influence of 'contamination', i.e., PSA testing that occurs outside the organised screening population should always be kept in mind. Perhaps most importantly, the results of all three peer-reviewed articles relating to the current study were reported *without* full consideration of the likely problems with their results with respect to length- or lead-time -biases. The results contained within these three articles were accepted for publication after peer review, but little, if any, direct mention was made by any of the peer-reviewers about the results being problematic with respect to length- or lead-time -biases. Such problems with the published results can be seen, at least in part, as a failure of the peer-review system to quality-control research output.

5.10 Assessment of economic evaluation methodology

One partial assessment of the methodology of economic evaluation is provided by Article III (Booth, 2019a), but the preceding sections of this study also aim to provide details of the main principles related to health economic evaluation. This sub-section summarises these descriptions and assessments with much brevity. Article III identifies three main approaches to health-economic evaluation namely, the extra-welfarist, welfarist, and classical. Table 1 is intended to provide a very brief description and summary of these three approaches (see (Booth, 2019a) for further details). Article III is based both on the author's reading of the literature and on experiences during the practical application of health-economic evaluation as part of this study. Discussion about such subjective assessment will continue in subsequent sub-sections, and it will be left for the reader to decide if she agrees with the interpretations drawn.

The first row of content in Table 1 describes welfarist approaches to health-economic evaluation (WHEE), and it is only this approach which ensures adherence to neo-classical (welfare-)economic principles. WHEE typically aggregates individuals' 'willingness to pay", using contingent valuation methods, in order to produce estimates of 'social welfare'. Its drawbacks include the vagaries of asking individuals to value goods or services about which they do not typically have much information (or experience). The second row of Table 1 describes extra-welfarist approaches to health-economic evaluation (EWHEE). However, this approach has little adherence to neo-classical economic principles, and instead typically relies on a combination of individuals' health-state -related assessments and some valuation of those assessments which is typically an amalgam of both individuals' preferences as well as the preferences or responsibilities of a 'decision-maker' (be the decision-maker hypothetical, mythical or other), rather than the preferences of individuals alone. For this reason, this often decision-analytic approach, with purported strong emphasis on a 'decision-maker', is sometimes referred to as the "decision-making approach". The third row of Table 1 describes a classical approach to health-economic evaluation (CHEE). This approach has the primary objective of identifying and assessing, rather than defining and 'maximising', societal value health' or well-being. This involves a process of analysis and evaluation which would include quality assessment and checking of appropriateness of the health-economic evaluation itself.

Table 1. Summary of the main approaches to health-economic evaluation found by the critical literature review.

Approach to evaluation:	Principles related to the approach:	Central objective(s) of the approach:	Important features of the approach in practice:
Welfarist (WHEE) (see, e.g., (Mishan, 1971) and (Birch & Gafni, 1996))	Analysis is based on neo-classical welfare-economic thinking (Dantzig, 1951) and individual preferences (Birch & Donaldson, 1987).	Tends to suggest that social welfare is some aggregation of '**individual welfare**', 'well-being' or 'utility' (somehow defined and measured) and should be '**optimised**' (Birch & Gafni, 2011).	Opportunity cost is typically defined as an integral part of WHEE, and will include those criteria and constraints which have been incorporated and about which there is sufficient information. (McDonald et al., 2016). WHEE suffers from measurement problems: uncertainty remains in many instances as to the nature of the preferences stated by respondents, i.e., there is often ambiguity as to exactly what is being measured by contingent valuation methods.
Extra-welfarist (EWHEE) (see, e.g., (Sugden & Williams, 1978), (Sculpher et al., 2005) and (Drummond et al., 2015))	Pragmatic assumptions are made that, e.g., the QALY approach does, to some extent, represent the 'value of health', such assumptions are combined with some form of constrained maximisation, in order to maximise 'output' from a given budget (Drummond & McGuire, 2001).	The aim is to '**maximise**' some form of (health-related) 'efficiency', typically using an indicator of '**health**', such as functional health status (Orr & Wolff, 2015). The chosen efficiency 'measure' is then compared to the efficiency of other ways to produce the same indicator of health or another standard of value, such as some kind of 'cost-effectiveness threshold'.	Crucially EWHEE is reliant on the available information concerning costs and effectiveness, as well as on the restrictions placed on the analysis by the chosen approach to EWHEE. If the standardised cost and effectiveness indicators used are not applicable to the topic in question, or if they cannot be applied for any reason, EWHEE may well *not* provide a competent assessment. Unless the resulting ICERs are compared to some ready estimate or diktat (i.e., so-called 'cost-effectiveness threshold'), then EWHEE is typically demanding for decision-makers as they must place a value on the EWHEE efficiency indicator.
Classical (CHEE) (see, e.g., (Franklin, 1842), (Quade, 1989), (Wildavsky, 1993) and (Coast, 2004))	Cost-benefit *thinking* based on the identifying and assessing the pros and cons of alternative courses of action, via deliberative analysis. Deliberative analysis being a process of evaluation which would include quality assessment and checking of appropriateness of the study itself, or of previous studies.	CHEE need not focus on a single maximand (as EWHEE typically does) or on a single source of preferences (as WHEE typically does). The aim is that '**societal value**' can be '**identified**' by politico-economic comparison of evidence and objectives, through multidisciplinary assessment and deliberation†	CHEE allows flexibility in the form of the analysis of value and opportunity costs. Relevant information relating to any analysis (formal or informal) would include: A) evidence about the relative **'**efficiency**'** with which 'health' or 'welfare' is being produced, B) evidence about budget or resource impact(s), as well as C) the quality of the evidence and, e.g., its coherence with the policy problem and its types and levels of uncertainty. This approach does not provide a systematic solution, but asks questions about the competency of the evidence on a case by case basis. The CHEE would describe the qualities of the available 'economic' evidence, or the lack of such evidence.

† also likely to include analysis, or information, resulting from the EWHEE and/or WHEE approaches.

6 Discussion

The main focus of the remainder of this thesis is on trying to convey some important features of the nature of information provided by health-economic evaluation. The review of the literature on health-economic evaluation undertaken as part of this study suggests information from CEA should be subjected to an evaluative process. The discussion, that follows on from earlier sections, attempts to draw together various arguments so as to make the case for a more detailed and deliberative approach to the interpretation of information produced by health-economic evaluation, especially that information which is produced via CEA. It will then put forward some suggestions which could be helpful when interpreting economic evaluations.

6.1 How should health-economic evaluation be interpreted?

The suggestion presented below is that interpretation of information from health-economic evaluations should first be assessed through a two-part quality-assessment process. The first part of the process would focus on the quality of the health-economic information produced by each health-economic evaluation, and primarily assess the *content* or 'internal' validity in each case (Sub-sections 6.1.2 and 6.1.3). A second part of the process would assess how that content fits into the *context* in question, i.e., it would appraise the appropriateness of health-economic evaluation information (Sub-section 6.1.4). Together, these two parts can be used to form an overall assessment of the 'quality' for each health-economic evaluation in any given setting (Booth et al., 2017) and contribute to any relative evaluation (as set out in Sub-section 6.2).

Before these two foci of preliminary interpretation can be set out more fully, it will be instructive to differentiate between 'measures' and 'indicators', in order to better discuss and understand the meaning and interpretation of much of the information provided by health-economic evaluations.

6.1.1 'Measures', indicators and meaning in CEA

Numbers can be assigned to act as measures under different rules, but not all of these rules are of equal usefulness in all settings (Porter, 1992). The numeric results from health-economic evaluation are typically assumed only to provide a partial contribution to holistic evaluation, they are intended to be an *aid* to decision making (Sculpher et al., 2005). CEA, in the field of health-economic evaluation, appears to be a rather undisciplined collection of methods (Bauer, 1992), which does not seem likely to produce wholly comparable or commensurate numbers. Indeed, rather than there being a lack of recommendations and guidelines suggesting how to undertake CEA, there are a plethora of different approaches (Heintz et al., 2016). Two related issues should also be noted here, even if one single standard approach did exist, firstly, there would likely be difficulties in following that standard approach for some applications, given that the information required in any standard approach might not be readily available for each and every application. Secondly, if one single standard approach existed, it would likely neglect to include potentially important elements for some applications. Given the lack of a systematic approach, given that the numbers would appear to be less than consistent in their meaning, it would seem inappropriate to class ICERs as a *measure* of efficiency, each ICER is merely an *indicator* of one form of efficiency (Stevens, 1959). Given the extensive lack of certitude surrounding ICER estimates it would seem even more appropriate to see the ICER estimates as merely indicators. As indicators, they can give some indication of some specific form of efficiency, or even some wider notion of efficiency or value, but their sensible use as indicators would seem to require some kind of quality control (Schum, 2009). Although these ICERs are numbers, they do not have consistency in meaning, they do not seem to adhere to the following meaning of 'measure': they do not seem to provide a standard, a rule of judgement, etc., against which something may be gauged, determined, or regulated (Oxford English Dictionary, 2016).

Although the existence of measurement or analytical standards are often assumed by many health economists (see, e.g., (Eastwood & Maynard, 1990)), other writers have been much more sceptical that about the extent to which objectivity can be attained (Goldenberg, 2006), stressing that such numbers should not be interpreted without careful consideration of what such information really means (Porter, 1995), including appropriate

qualifications as and when necessary (see, e.g., (Boulding, 1966)). This is not to say that standardisation, recommendations and guidelines have no role, but rather that when these quasi-systematic approaches are utilised, practitioners should clearly acknowledge that CEA is typically truncated, and may exclude important wider considerations. Further, the diversity in estimates of 'health' (Maynard & Sheldon, 1997), the diversity in estimates of the production of 'health' (Lessard & Birch, 2010), and in the diversity in the synthetic indicators of the *efficiency* of producing 'health' (Brousselle & Lessard, 2011), would seem to make them too heterogeneous to be used in an unquestioning manner as a metric (Muller, 2018). Given that information produced by health-economic evaluations is often disparate in nature, it is suggested here that the resultant numbers might primarily be classified towards the nominal end of the measurement-scale spectrum (Porta, 2014, p. 180). Indeed, the diversity in estimates of the efficiency of producing 'health', i.e., the inherent variability in CEA estimation processes, would seem to provide another reason why evaluation of what such information means on a case by case basis, including judging its provenance, would seem useful (Laine, 2014).

At least two further related reasons exist for quality assessment of CEA information, i.e., for deliberating upon, rather than relying on assumptions about, what CEA means. The first is that, buoyed by policy-oriented rhetoric (Colvin, 1987), information from health-economic evaluation has long held the promise of being an administrative aid (Feldstein, 1963), despite its limitations (Walsh & Williams, 1969). Second, the development of certain streams of cost-benefit thinking for policy purposes was likely a long-standing synergistic relationship between industry and government, via research institutions, both public (see, e.g., (Mason & Towse, 2008), (Cookson & Claxton, 2012) & (Cookson et al., 2016)) and private (see, e.g., (Laing, 1972), (Office of Health Economics, 1979) & (Brockis et al., 2016)).

Although these two latter reasons are not within the primary focus of the current study, such sociological or philosophical considerations should nevertheless be kept in mind when assessing all evaluative research (see, e.g., (Boulding, 1966), (Biderman, 1966), (Ashmore et al., 1989), (Valtonen, 1993, p. 193), (Lessard & Birch, 2010), (Howick, 2011) and (Dempsey, 2019)). Indeed, mesmerised by the 'beauty' and precision of numbers, decision-makers may easily overlook the simplifications made to achieve such (oftentimes spurious) precision in ICER estimates. Neglecting analysis of

qualitative factors may overemphasise the importance of abstract or ideologically -based calculations to the decision process (Quade, 1967). Given all the reasons set out above, appraising the evaluation mechanism *per se*, e.g., its promises, pitfalls and possible biases, would seem to constitute a necessary and integral part of any comprehensive evaluation of the meaning of information from health-economic evaluations. For this reason, we now turn to the first part of a two-part appraisal process, which should involve the health-economic evaluation being checked for 'content'.

6.1.2 Appraising the *content* of CEA information

The previous sub-section on measurement attempted to highlight the need to define meaning for information from each health-economic evaluation. Much of this type of information does not naturally lend itself to investigation using standard terms such as reliability, reproducibility, validity, feasibility and sensitivity (Kobelt, 2013), due to a lack of relevant standards against which these concepts can be judged in the case of health-economic evaluation. However, it is the contention of this thesis that an attempt should be made to apply some kind of appraisal to economic evaluation *per se*, given that economic evaluation is made up of estimated outcomes and costs, as well as combined with assumptions or some form of analysis, which produce the particular efficiency indicator in question. In order to distinguish between the appraisal of the content of health-economic information and the appraisal of that information in any specific context, the terms 'credibility' and 'appropriateness' will be used, respectively. The use of the term 'credibility' here is intended to express a broad notion, akin to concepts such as 'construct validity', 'internal validity' and 'plausibility' found in the epidemiological and health-economic literatures. The term credibility will be used to refer to an assessment of the components which form the constituent parts of the CEA, including both the estimates of costs and effectiveness, as well as the manner in which they are combined.

Subsequently, given the credibility or plausibility of that quality-assessed information, its congruence with, or appropriateness to, the policy question should also be assessed. In practice this would mean that, subsequent to checks for credibility, attention should then turn to the bigger picture; looking at how CEA information fits in with other evidence, i.e., if and how appropriate 'economic' evidence would seem to be after being qualified by any other

considerations. As the topic to which economic evaluation is applied gets more complicated or complex (Shiell et al., 2008), we should be more sceptical about the possibility for economic evaluation to provide useful information (Kurtz & Snowden, 2003). This does not necessarily mean that no useful information will be available, just that care should be taken to challenge anyone who suggests CEA is necessarily anything more than sub-optimisation for more complex or complicated topics.

The fundamental problem here is that health interventions can be horrendously complicated to assess (Williams, 2004, p. 14). Given the inevitable gap between the research that can be offered and what would ideally be available for decision-making processes, the suggestion is made that researchers would take heed of this and not assume that CEA and related methods could be legitimately applied to all topics with equal success. Researchers should be wary of the fallacy of composition; that what may 'work' in one area of health care may not 'work' in all areas of health care. There is no dispute that CEA can work in principle, but here the focus is on practical application of health-economic evaluation to health-care interventions, sometimes in complex social settings.

That said, appraising the content of the cost- and effectiveness -indicators which have been measured or estimated would typically start by checking what entities were measured or estimated and in what way. For example, how has the information been derived (e.g., study type)?; duration: months, years, or other?; what has been modelled using what information?; is the follow-up period sufficient to encompass all important impacts?, *et cetera.*

The main problem is that quality seems unable to be assessed in anything but a subjective or superficial manner via any of a number of 'checklists' (Evers et al., 2005; Husereau et al., 2013; Shemilt et al., 2013; Shemilt et al., 2008) and (Drummond & Jefferson, 1996). Although these checklists are available for assessing the 'methodological quality' of, or the standard of reporting of, health-economic evaluations, they do *not* necessarily address the *quality* of health-economic evaluations *per se.* The study by Evers et al. suggests that answering checklist questions would tend to require personal opinion and judgment, and that the questions they contain would tend to be quite cursory (Evers et al., 2005). However, the scope of the current study does not include the creation of a more, in depth or detailed, credibility assessment instrument, presuming such an instrument would be sensible, or even possible, to construct.

As an alternative, appraisal of the CEA results for content could be facilitated to some extent by the use of an impact inventory for CEA, by appropriately appraising each source of evidence for which an impact category has been assessed (see Table 2 below (as suggested by (Neumann & Sanders, 2017))).

6.1.3 Credibility and the Impact Inventory

As suggested in the previous sub-section, some of the types of impacts which could usefully be checked when considering 'credibility' are set out in the table below, this template for this table was developed by the Second Panel on Cost-Effectiveness in Health and Medicine (Neumann et al., 2016, pp. 351-352) and is filled in here according to the information available from Article IV:

Table 2. Potential impacts related to CEA quality assessment

Sector	Type of Impact (list category within each sector with unit of measure if relevant)†	Included in this Reference Case Analysis from the so-called '… perspective':		Notes on sources of evidence:
		Healthcare sector	Societal	
Formal healthcare sector				
Health	**Health outcomes (effects)**			
	Longevity effects	☑	☐	
	Health-related quality-of-life effects	☐	☐	
	Other health effects (e.g., adverse events and secondary transmissions of infections)	☐	☐	
	Medical costs			
	Paid for by third-party payers	☑	☐	
	Paid for by patients out-of-pocket	☐	☐	
	Future related medical costs (payers and patients)	☐	☐	
	Future unrelated medical costs (payers and patients)	☐	☐	
Informal healthcare sector				
Health	Patient-time costs	NA	☐	
	Unpaid caregiver-time costs	NA	☐	
	Transportation costs	NA	☐	
Non−healthcare sectors (with examples of possible items)				
Productivity	Labour market earnings lost	NA	☐	
	Cost of unpaid lost productivity due to illness	NA	☐	
	Cost of uncompensated household production††	NA	☐	
Consumption	Future consumption unrelated to health	NA	☐	
Social services	Cost of social services as part of intervention	NA	☐	
Legal or criminal justice	Number of crimes related to intervention	NA	☐	
	Cost of crimes related to intervention	NA	☐	
Education	Impact of intervention on educational achievement of population	NA	☐	
Housing	Cost of intervention on home improvements (e.g., removing lead paint)	NA	☐	
Environment	Production of toxic waste pollution by intervention	NA	☐	
Other (specify)	Other impacts	NA	☐	

† = Categories listed are intended as examples for analysts, NA = Not applicable, †† = Examples include activities such as food preparation, cooking, and clean up in the household; household management; shopping; obtaining services; and travel related to household activity.

One point of note about any use of an 'Impact Inventory' such as that depicted in Table 2 is its ability to highlight all direct or indirect impacts, despite the fact that, in practice, many CEAs have not included health effects on persons other than the individuals directly affected by the interventions (Neumann et al., 2016, p. 92).

One possible drawback of the Impact Inventory, as presented by Neumann et al., is the fact that it does not separate out all health impacts in as much detail as it does with cost impacts: it does not appear to leave much scope for including health effects under the heading of "Informal health care sector". Of course, any health impacts which fall on 'important others', i.e., on relatives or carers, could be bundled under the subheading "Other (specify)", itself under the heading "Non-health care sector". However, all other items in this section seem to refer mainly to costs, rather than to health *per se*. Further, the table could be extended to encompass such concerns and, thus potentially provide an improved vehicle for assessment of 'credibility'. Naturally these issues will also have importance for the consideration outlined in the next section, namely, 'appropriateness', too.

6.1.4 Appropriateness: CEA information in *context*

Assessing 'appropriateness' means assessing the practical relevance of the CEA information produced in informing choices. Of course, any assessment of appropriateness will naturally include both the relevance of the constituent measures used, as well as the method used to combine both costs and effects, as set out in the previous sub-section. In addition to the above initial assessment of 'credibility', the appropriateness of the CEA indicator to the policy question should be judged. Assessing the appropriateness of CEA information requires assessment of the extent to which the available research evidence meets the needs and concerns of decision-makers. It should be recognised that CEA information can vary in both its credibility and impartiality, as well as keeping in mind that the production of other forms of evidence in the health-care sector is not without its problems either (Gøtzsche, 2019). The extent of the appropriateness of the 'decision problem' to the 'policy question' has central relevance for the way in which CEA informs decision making. To put this another way, any approach chosen when undertaking CEA is only one of a myriad of possible approaches: appropriateness is not

guaranteed, it needs to be earned and assessed. Ideally, assessing the appropriateness would also check how far the results of an iterative evaluative process, e.g., CEA, remains from the policy question itself (see Figure 1., sub-section 5.4). In part, appraisal of the CEA results in any specific context can, just as when appraising credibility, be facilitated by the use of an impact inventory for CEA (see Table 2, sub-section 6.1.3). Appraising the appropriateness would typically start by checking what indicators of cost, effectiveness and cost-effectiveness, i.e., what impacts are included (Loomes & McKenzie, 1989). In addition, of crucial importance is *from whose perspective(s)* are these entities assessed: patient?; health-care payer?; cross-government?; health-economic?; family?; community?, or a mix of these.

Appraisal may also touch on both content and context in the following categories: the measures or indicators used, the data used, and study design(s) in question. Such appraisal, requires judgment about how well the data, the records or responses, describe the potential benefits and harms, i.e., how responsive they are (see, e.g., (Strech & Tilburt, 2008), (Furlan et al., 2015) or (Kobelt, 2013)). In part this appraisal must uncover to what extent each CEA *in practice* matches, or fails to match, ideal 'utilitarian' evaluation (Brock et al., 2016). Measurement problems, especially surrounding those important considerations which are simply not amenable to measurement (see, e.g., (Neumann et al., 2016) or (Feeny et al., 2016)), have always been around to thwart the utilitarian project (Jonsen, 1986). Given the above, three key considerations which should be included when assessing the appropriateness of a CEA to a policy question would typically be:

1. What potentially important costs have been omitted?
2. What potentially important effects have been omitted?
3. Are the methods used to combine the costs and effects appropriate?, especially in the particular decision context in question.

As noted before, answers to these questions would require judgments to be made about how appropriately each CEA describes the likely relationship(s) between 'important' costs and effects of the intervention, be they measureable or intangible.

6.2 Two main ways to interpret CEA

Armed with a preliminary interpretation of both the credibility of the CEA information *per se* and how the content of that information fits in the context of the holistic policy question (its appropriateness), there is now a need for a third aspect of interpretation: relative valuation. As noted above, CEA estimates are conditional on the methods used and on how successfully those methods have been applied, and appropriateness to the policy question of those estimates is case specific.

In this third aspect of interpretation, namely relative valuation, judgment is also required relating to the term 'cost-effectiveness'. Two approaches to this will be set out here. First, the dominant extra-welfarist approaches will be described, this is one which combines ICERs and a threshold (in the next sub-section). Subsequently, the rationale for a second approach will be explained, this 'classical' approach combines, e.g., ICERs with information related to budget impact through deliberative processes (in sub-section 6.2.3). Before that, it will be useful to clarify the terms 'cost-effectiveness' and 'cost-effective', in order to help understand both of the above approaches to relative valuation.

In the context of interpreting health-economic evaluation, statements are often made which claim that "intervention A is more cost-effective than intervention B". Although even this relative statement is demanding in terms of its informational requirements, it merely expresses a comparison of the two interventions in relation to each other, i.e., a simple judgment concerning an ordinal relationship. The informational requirements referred to above would be that the measures or indicators underlying the CEA are of high enough quality, and are sufficiently apposite to the policy question, to feel able to draw a conclusion about intervention A being *more* cost-effective than intervention B. This expresses a relative, not absolute, valuation. On the other hand, a claim that "intervention A *is* cost-effective" is typically interpreted an absolute statement and, in addition to being demanding in terms of its informational requirements, requires a reference point. Any such claim is a double-edged sword, it inherently requires the use of some kind of 'threshold', but it also produces a statement which has some level of allocative or normative intent. Thus, in order to make statements absolute about 'cost-effectiveness' requires explicit or implicit value judgements to be made, but also produces a value-laden conclusion. Two of the main approaches taken to these value judgments will be outlined in the next two sub-sections.

6.2.1 'Extra-welfarist' interpretation of CEA

Applications of the extra-welfarist approach generally utilise incremental cost-effectiveness ratios (ICERs) in the form of mean cost-per-QALY estimates, with a notion of opportunity cost of the additional 'health' production also being expressed as some threshold, i.e., a cost-effectiveness-ratio threshold (CERT), somehow defined. The qualification 'somehow defined' can be seen as very important as although extra-welfarist approaches *per se* do not specify any particular maximand, any threshold, or any comparative standard, would ideally be defined in exactly the same terms. A less common approach, which ostensibly compares estimates of ICERs for similarly-assessed technologies, employs a 'league table', i.e., a ranking of comparator interventions (Williams, 1985). When considering the more common threshold approach (Culyer et al., 2007), it should be remembered that the QALY approach is itself a concept that is applied using many different methods, with a variety quality-adjustment 'measures' available, i.e., different health-related quality-of-life indicators capturing any of a range of different dimensions of health status and its valuation (Richardson et al., 2016). It should also be noted here that similar variations in methods affects the estimation of average changes in the number of life years which is also part of the QALY 'metric'. To summarise the above, the claim is made here that the QALY is not really a measure, it is more a concept and an approach, i.e., an indicator of quality of life over a specified period.

In a similar manner, although health status is usually assigned a number or weight, there are many different processes and rules under which such estimates or indicators of health status are produced (Booth et al., 2014a) (either in a generic or specific sense (Brommels & Sintonen, 2001)). Further, QALY estimates, however they are derived, in practice are typically underpinned by a RCT-based, bio-clinical model of inputs and outputs relating to the disease in question (Garrison et al., 2018), even this approach has its weaknesses. For example, RCTs can provide information on the average treatment effect in terms of effects and costs, but in the presence of outliers they are likely to be deficient (Deaton & Cartwright, 2018).

Although it is quite widely thought that QALY approaches can, in many circumstances, provide a useful representation of the value of health status, it often only provides an inherently partial estimate of changes in '*health*' (Loomes & McKenzie, 1989). To be more precise the QALY approach typically

only provides an inherently partial estimate of changes in *'some health-status-measure -proxy for a quasi-societal valuation of health'*. The practical relevance of this is that EWHEE, rather than promoting maximum 'health' gain, typically only maximises QALYs, which tends not to maximise health *per se* (Mooney, 1989). Despite these limitations, a variety of manifestations of the QALY framework are often proffered as the 'best available' approach for the (quasi-)systematic, quantification of a generic concept of 'health' (see, e.g., (Garrison et al., 2018), (Thokala et al., 2018) or (Neumann et al., 2016)).

The suggestion is made that, rather than *relying* on any such 'best available' extra-welfarist CEA approach, even merely as a starting point for discussion, its every application should be subjected to appropriate quality assessment (in line with Sub-section 6.1). The extra-welfarist CEA approach promises much, but can fail to deliver on those promises, one way in which this failure can occur is the focus of the next sub-section; i.e., on the extra-welfarist approach which compares mean cost-per-QALY in relation to some threshold value (McCabe et al., 2008).

6.2.2 Critique of the current state of the art

The relevance of extra-welfarist approaches to policy-making processes is limited in both principle and practice. One *practical* criticism of extra-welfarist approaches is the lack of commensurability between CEAs and, although this problem may be mitigated by the use of a reference case, reference cases are only guidelines and not always practicable prescriptive rules (see, e.g., (National Institute for Health and Care Excellence, 2013) and (Neumann & Sanders, 2017)). Indeed, as commensurability is diminished by natural, understandable and legitimate variations between CEAs, this would seem to provide one reason why a single CERT, i.e., a threshold or threshold range for cost-effectiveness, would not seem to be a wholly justifiable standard in practice; either within a single organisation (e.g., NICE (National Institute for Health and Care Excellence, 2017)), or even within a single programme (e.g., NICE's single technology appraisal (STA) process). Indeed, the current absence of unified guidelines for health-economic evaluation would seem to confirm the earlier suggestions that, *a priori*, it is unwise to determine which type of study should be performed (Glick et al., 1994; McIntosh et al., 1999; Strech & Tilburt, 2008). The rather arbitrary and *ad hoc* nature of applied health-economic evaluation

seems abundantly clear, but, on the basis of the literature review undertaken for the current study, seems rarely to have been acknowledged by many of its staunchest supporters (Drummond et al., 2015). That said, that hard line might now be softening, with one commentator increasingly making reference to health-economic evaluation as a skilled art (Culyer, 2014, p. 438; Culyer, 2018), rather than his previous comments describing it as an applied science (see, e.g., (Culyer, 1976, p. viii)). A second reason for doubting the *practical* usefulness of a threshold stems from the fact that a threshold relates to extra-welfarist *principles*, the idea of a threshold is little more than a postulate and seems only adequate for something akin to abstract Pigouvian theorising (Pigou, 1947, p. 31). Such marginalist ideas tend to detach themselves from discussions about value, instead 'value' in marginalist theory is assumed to be encompassed by price (Klarman, 1965, p. 171). These price-based marginalist theories of 'value' seem to be debunked, on the one hand due to the practical heterogeneity of CEA methods and, on the other, by the principled need for deliberation and judgment about what society values within decision processes. Indeed, weaknesses in the notion of the cost-per-QALY threshold can also be a seen as weaknesses in the notion of the label 'value for money', often used by some health economists. This is because, given the absence of a market and the absence of a threshold price, CEA cannot provide more than partial information regarding 'value for money', as alluded to in the following quote:

> "It must ultimately be recognized that only within rather narrow limits can human conduct be interpreted as the creation of values of such definiteness and stability that they can serve as scientific data, that life is fundamentally an exploration in the field of values itself and not a mere matter of producing given values. When this is clearly seen, it will be apparent why so much discussion of social efficiency has been so futile." (Knight, 1923)

As already noted in sub-section 4.3, due to a dearth of widely-accepted quality-assessment guidelines for CEA, it typically remains a matter of subjective judgment as to what extent any CEA successfully provides robust evidence. The estimates of costs, effects and cost-effectiveness are not immutable facts (Williams, 1974). Economic evaluation is most often not based on wholly systematic, wholly reproducible, or wholly objective scientific research, it is merely partial evaluation: in each application of health-economic evaluation the economic value framework in question *per se* provides the value-laden framework for valuation.

One succinct way to convey some of the above reasoning is via a cartoon, as in Figure 6, below. This cartoon can, of course, be interpreted in many ways, but it does seem to skilfully and aesthetically encapsulate the idea of a 'market' for QALYs.

Figure 6. A satirical cartoon about health economics and planning (Paddison, 1986) (reprinted with the permission of the artist)

In line with the rest of the current study, a number of interpretations are suggested for this cartoon. It should immediately alert us to the relative inertia of CERT estimates: the mythical figure of £30,000 per QALY is still one of the most widely quoted in the literature even today, and many health economists feel that fact alone should raise alarm bells (Caro et al., 2010).

Figure 6 could also be interpreted to reveal the power of voice, as described by the 'all-singing, all-dancing' -members of "Social Class One", as well as the ubiquitous presence and power of civil servants or decision-makers. It nicely illustrates that the use of an approach which includes CERTs would necessarily include a need for judgment on the part of any single decision maker and, fittingly, in this figure there are many. Further, Figure 6, with the

brevity only art can muster, describes the lack of reality of a *fixed* CERT (Coast, 2017): this is a sale room. In a sale room, the idea that "a QALY is a QALY is a QALY" is soundly rejected, a myriad of factors affect how those on display perform, as well as how bidders react. Thus indicating that, rather than there being an orthodox, neoclassical, market-clearing price, the valuation of QALYs is better depicted as an auction. In an auction, value will, in part, always be 'in the eye of the beholder' (the bidder), in part being tempered to some extent by the other bidders in what could be described as a quasi-market. An alternative interpretation would be that the saleroom itself (the system of economic evaluation itself) has a purpose, it exists in order to serve the needs of the vendors (the most powerful of whom in this case could, e.g., be the pharmaceutical industry). However, even if we do not agree with the assumptions necessary to create a price-hiking neoclassical economic market, i.e., the central tenet of EWHEE approaches, a functioning market, seems nonsensical (Self, 1970).

A fitting paraphrase might now be that "QALYs are not QALYs are not QALYs": they *are* typically dependent on who gets them, as represented in the cartoon by the many and varied potential recipients of priority. Williams also offered the following challenge (in a paper published posthumously):

> "The assertion is that, from a public policy perspective we assume that the value of a QALY is the same no matter who gets it. Anyone is free to challenge that assumption and propose another, and we could then examine the differential consequences for priority setting. But some assumption has to be made and justified." (Williams, 2005)

In response to the challenge Williams laid down, it no longer seems sustainable to rely on a fixed cost-per-QALY threshold for heterogeneous impacts on difficult-to-measure 'health' (Booth, 2019a). The current study strongly suggests that value is not confined to QALYs, higher-order values and value definitely exist, and would seem to require both respect and recognition. A more nuanced approach seems necessary as assuming that the value of a QALY is the same no matter who receives it, or assuming that the QALY encapsulates all dimensions of value, is simply too abstract for uncritical use in robust planning. A suggested alternative approach will be outlined in the next sub-section, in line with the idea that a range of different evaluative approaches will likely be useful in order to encapsulate these higher-order values. EWHEE has a tendency to focus on its own efficiency indicators, potentially to the detriment of other considerations. Examples include a

tendency to ignore any budgetary impact of the efficiency measures it produces (Birch & Gafni, 2003; Garrison et al., 2018), as well as its tendency to focus on the magnitude of any ICER, even when the absolute size of the health impact may carry more weight with many stakeholders (Kelleher, 2014).

In addition, the phrase 'decision rule', which has a strict meaning within abstract decision-analytical approaches, does not necessarily have much meaning outside the EWHEE approach in question. Such potentially misleading EWHEE terminology may also contribute to the crowding out of other values from prioritisation processes, by raising expectations that the EWHEE approach could have a direct impact on the decision *per se* (Booth, 2019b). To reiterate, the output of these neo-classical economic or management science -solutions do not typically offer holistic evaluation, they are too partial; they do not normally take into account effects as disparate as patient satisfaction, medical-industry profits, or environmental sustainability.

Orthodox economics makes an abstraction, simplifies choice problems by making them peculiarly narrow, perhaps mainly in order to make them amenable to the decision-analytic approaches of management science. Simple notions of economic efficiency can cause the neglect of higher-level efficiency, i.e., the second and higher-order effects of policies, e.g., those related to system sustainability or community values. In summary, the main critique of the current 'state of the art' in extra-welfarist health-economic evaluation is that it does not do enough in distinguishing between some of the abstract principles related to health-economic evaluation and the information it can provide to prioritisation processes (Booth, 2019a). This failure likely results in an increased likelihood that we forget the important opportunity costs of taking the wider range of into account.

To largely ignore many of the above issues, as some extra-welfarist economists sometimes seem to do (Green, 1990), in rhetoric or in practice, seems arrogant. For that reason if no other, an alternative assumption will be proposed in the next sub-section to help return to the foundations of HTA (Daddario, 1968). This proposal aims to supplement the often simplistic and flawed approaches to measurement used by neo-classical economics, in order to better take into account the second and higher-order effects of policies.

6.2.3 A more classical approach to interpreting CEA

Given the critique presented in the earlier parts of the current study, and especially those outlined in the previous sub-section, the use of approaches relying on a single threshold (or threshold range), or on contingent valuation methods, are now taken to be soundly rejected as comprehensive solutions. The rejection of these mainstream neo-classical economic approaches in part stems from the inherent conflict between restricting the scope of measurement to a core set of dimensions (in order to facilitate comparability and commensurability) and the aim of producing comprehensive valuation. Because the measures available cannot always be successfully applied, CEA information typically remains incomplete, i.e., truncated and partial analysis (see, e.g., sub-sections 4.3, 4.4 or 5.4, or (Quade, 1967)). Quantitative analysis is also limited by the presence of intangibles, i.e., entities which cannot be measured, or those that cannot be accurately or precisely measured. In order to understand what each application of CEA means, it is crucial to be able understand what is included, what is left out from almost all of its many forms: CEA usually misses out essential dimensions relevant to many prioritisation processes (see, e.g., (Quade, 1967), (Wiseman, 1989a) and Article III). To recap, any decision-maker must decide the relative worth of the efficiency indicator as estimated in the study, and is under no constraint, logical or otherwise, to apply a fixed value to, e.g., QALYs in different settings (Evans, 1984, p. 262).

For these reasons a more 'classical' approach to interpreting CEA is proposed below: in line with one original aim of cost-benefit thinking (Franklin, 1772), of course CEA does indicate a form of economic efficiency, but potentially that is all it indicates, it does not, on its own, necessarily indicate 'worth' or 'value for money'. Therefore, the claim is made here that the most appropriate interpretation of information typically provided by CEA is that it provides an unreliable indicator of some form of economic efficiency. Broader interpretations, i.e., those which could be more closely related to 'worth' or 'value for money' would seem better served by a process of judgment, as part of what the current study calls classical HEE (CHEE). Under CHEE, value is defined by a deliberative process, by deliberative analysis of all the available information, rather than by reference to any 'threshold' or any associated 'decision rule' under EWHEE, or by accepting the results of, e.g., the contingent valuation approach under WHEE. In short, CHEE can be described

as a responsive approach, rather than the rather pre-defined frameworks typical in EWHEE or WHEE. For example, consideration of the extent to which a technology transfers resources from consumers or taxpayers to the producer of a technology, i.e., the level of industry profits which necessary, normal, optimal or fair may be warranted (Brouwer et al., 2019)

6.3 CEA in relation to holistic evaluation

6.3.1 What can deliberative analysis add?

Previous sub-sections have provided a number of reasons why, in contrast to the view of (Drummond et al., 2015, p. 314), health-economic evaluation does not provide a thorough assessment of 'value for money'. Instead, this sub-section attempts to outline in what ways EWHEE and WHEE typically address elements of related policy questions and, just as importantly, fail to address 'efficiency' in a wider sense (Wildavsky, 1967) as noted in sub-section 5.10. This outline also proposes the use of CHEE and tries to explain and promote that proposition using an 'economic value plane', i.e., a representation of three dimensions of economic value, in Figure 7.

In the figure below, some of the reviewed literature on HEE and prioritisation is drawn together (e.g., (Norwegian Ministry of Health and Care Services, 2017), (Booth et al., 2017) and (Liliemark et al., 2016)). This figure depicts three dimensions along which different approaches to HEE contribute varying amounts of information about 'economic value' to prioritisation processes. Depending on the jurisdiction or on the technology in question, prioritisation processes may stress differently the importance of (1) information about need in terms of the 'efficiency' of the technology (in the form of a programme's ICER), (2) information about resource use (in the form of a programme's budget impact), and (3) information about other aspects of 'economic value' (including both appraisal of the qualities of any formal analysis in dimensions (1) and (2), as well as interpretation of any other evidence or values in dimension (3), in line with Sub-section 6.1.). It should also be noted that, from a strict welfarist perspective, economic efficiency cannot be established independently of the budget impact, since budget impact determines the resource requirements to be found from elsewhere and hence the benefits to be forgone. As a result these two

economic elements are not separate dimensions from a WHEE perspective. Despite this, it seems instructive to see the potential interplay between likely important dimensions of 'economic value' in Figure 7. For instance using CHEE, a mean ICER estimate for a technology could be relatively high (even after the quality of the evidence for the estimated mean ICER has been independently assessed) and the estimated budget impact of a technology could also be large. However, even in such a case, under CHEE, if there are other important 'economic' considerations with a positive impact on value, then this technology could still achieve a relatively high point on this 'economic value plane' (as marked by the red ellipse, with points towards the North-East typically preferable to those in the South-West). In this hypothetical example there are also two other small, coloured ellipses, the magenta ellipse is gives an example of the 'economic value' conferred via EWHEE and the navy blue one an example of the 'economic value' conferred via WHEE. Together these three ellipses show that different economic value will be placed on the same technology, by the different approaches to HEE (as set out in Article III), in part due to the extent to which they provide information on the three dimensions presented. Wider notions of value, from other non-economic viewpoints or from other disciplines may, of course, further modify their position after this analysis of 'economic value' (Dobrow et al., 2004). It is also acknowledged that post-HEE deliberation can take place under any approach to HEE, and as this could result in all three approaches ending up at very near the same value point on a 'decision-making value plane' (not depicted here). Beyond this 'economic value plane', dependent on the types of health-economic evaluation sought and obtained, there are still obvious opportunities for other values to enter any appraisal process. Two significant points should be noted, first that CHEE is unlikely to be anything like an exact science (Baltussen et al., 2017; Marsh et al., 2017), and second that typically each health-economic evaluation will tend to produce one single-shot CEA (even though this will be often include sensitivity analyses), i.e., one description of 'economic' value.

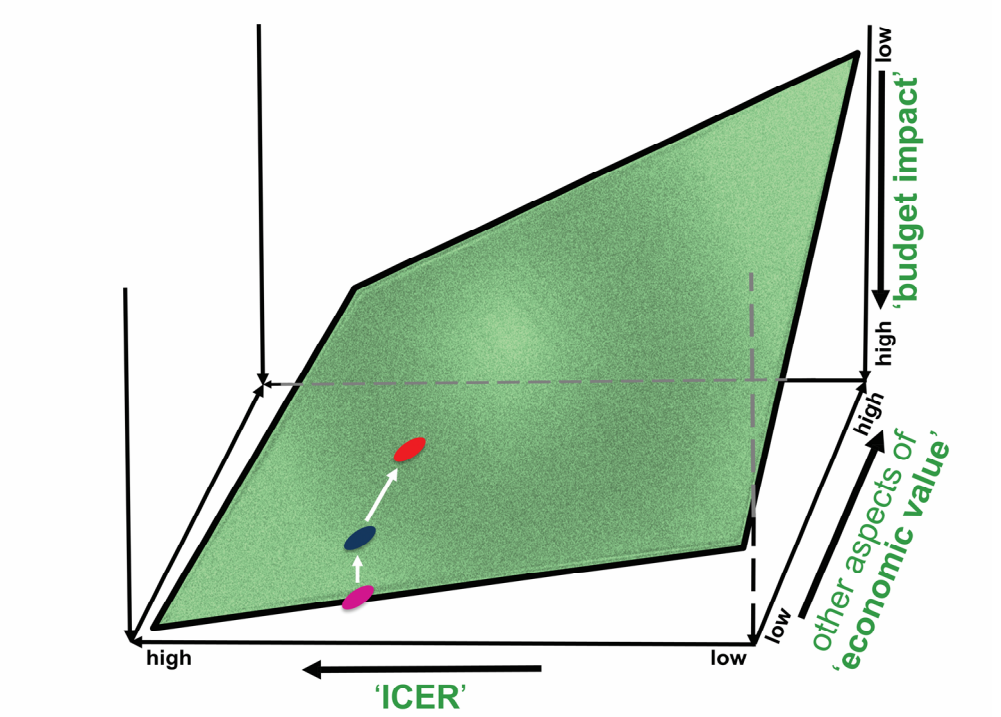

Figure 7. An 'economic' value plane

Other aspects of need, assessed in a quantitative manner could be incorporated into the third dimension of this figure (Klarman, 1965, p. 172). However, as Figure 7 represents an *economic* value plane, if qualitative or deliberative evidence is included in the chosen approach to economic value, a situation could result where 'economic' value becomes indistinguishable from holistic value. If such a point is reached, then it could be claimed that a 'cost-benefit thinking' -ideal would have had been reached, too. Economic evaluation has been transformed into valuation *per se*. This plane is intended as an informative descriptive device, a helpful representation, to try and help users of information from health-economic evaluations think about what different forms of health-economic evaluation might offer in terms of a holistic interpretation of value, a point to which the thesis will now turn for the CEA of the FinRSPC.

6.3.2 Interpretation of the FinRSPC's CEA information

In this sub-section some of the qualities of the components of the CEA of the FinRSPC in relation to the broad policy question mentioned in Section 3: "Can economic evaluation help us to know if prostate-cancer screening is 'worthwhile'?", are briefly described. However, when looking at the following 'results', it should also be kept firmly in mind, e.g., that *this* health-economic evaluation of the FinRSPC is an evaluation of *historical* effectiveness and practices, *in a particular Finnish setting*. All of the results here are therefore truncated temporally and in terms of both generalisability or scope.

As can be seen from Table 3, below, an attempt will be made not oversell the results of the health-economic evaluation of the FinRSPC. Table 3 shows that, even at a level of granularity suitable to avoid excessive cognitive burden (Miller, 1956), numerous questions remain to be answered before being able to make informed judgments related to the pros and cons of prostate-cancer screening (Steineck et al., 2019). Table 3 is not meant to be a once-and-for-all answer as to the 'economic value' of prostate-cancer screening, rather it is meant to be a work-in-progress and updated iteratively in accordance with Figure 1, in Sub-section 5.4.

Table 3. A preliminary set of results from cost-benefit thinking about prostate-cancer screening, on the basis of the FinRSPC after 20 years of the trial.

Pros of prostate-cancer screening	**Cons of prostate-cancer screening**
Men saved from prostate cancer death, on average**?**	Men overdiagnosed with prostate cancer**?**
Lives of some men lengthened, on average**?**	Lives of some men shortened, on average**?**
Positive HRQoL impact in men treated, on average**?** (men (over)treated for indolent cancers may rate up their HRQoL despite experiencing negative side effects)	An increased need for incontinence pads or pants, on average**?**
Decreased costs for some groups of men, on average**?**	Increased costs for some groups of men, on average**?**
…this list could be supplemented and updated as more information becomes available, and as deliberative analysis proceeds…	…this list could be supplemented and updated as more information becomes available, and as deliberative analysis proceeds…

As noted in Table 3, cost-benefit thinking would be supplemented and updated through a deliberative process as more robust and relevant information becomes available. The uncertainty described in Table 3 can be seen as debilitating, or as empowering, but it is the contention of this thesis that the lack of relevant information and the lack of ability to obtain appropriate information should be acknowledged (Kay & King, 2020). It seems useful to know that this health-economic evaluation of the FinRSPC contains few robust results, and clearly only provides a very partial answer to the tentative policy question: "Can economic evaluation help us to know if prostate-cancer screening is 'worthwhile'?". However, one tentative response can be given, that health-economic evaluation helps in revealing the *lack* of knowledge about whether or not "prostate-cancer screening is 'worthwhile'?". This in itself may be an important finding, and is supported by the suggestion that *incredible* certitude should be avoided (Manski, 2019).

Further, the FinRSPC-study results presented here cannot offer any prescriptive conclusion or policy advice, other than that it casts more than a little doubt on some aspects of earlier modelling studies. The suggestion of this thesis is that information from CEA can be usefully supplemented by other information, including health-economic information, is not new (Brousselle & Lessard, 2011; Coast, 2004). However, the suggestion that the qualities of health-economic evaluation be assessed in relation to what might be expected by decision-makers from more holistic evaluation, i.e., that deliberative analysis be used, appears to be a rarely-raised one in the current literature on health-economic evaluation. Despite challenging entrenched interests, the contribution of highlighting misleading language and practices in dominant forms of health-economic evaluation is a relatively simple one, and its worth will be left for the reader to decide.

7 Conclusions

The primary aim of the current study is not to denigrate economic evaluation, but to make its scope clear and to discuss its problems, limitations and potentials. The aim is to promote deeper understanding of the information which economic evaluation can provide. First, conclusions will be made about what it would be useful to understand in general about economic evaluation. Namely that, given the terminology of health-economic evaluation, the frequent deficiencies in the data or assumptions underlying analysis and the intentionally restricted scope of some approaches to economic evaluation, there will typically be many benefits in extending and checking the information provided by many approaches to health-economic evaluation.

7.1 What should be known about economic evaluation?

Economic evaluation tends to focus on measures of efficiency which do not encompass all relevant impacts on health, well-being or costs. Further, approaches to economic evaluation are usually forced to make strong assumptions in the estimation of costs and benefits. The resulting numerical representations should only be treated as meaningful when such meaning is warranted. In principle at least, this places demands on anyone producing or using cost-effectiveness analysis (CEA), as although efficiency can be interpreted as a simple concept, when applied to health care it is fundamentally complicated by measurement problems. These measurement problems include the wholly intangible, as well as the partially intangible: such as impacts which may not be captured satisfactorily when undertaking health-related quality-of-life measurement using generic health-related quality-of-life measures. In order to understand if efficiency indicators are misleading, disingenuous, or worse, economic evaluations need to be checked.

7.1.1 Efficiency is protean and needs to be checked

Information about efficiency varies and, therefore, needs to be judged in light of its credibility and appropriateness: it needs to be checked. Without the appraisal suggested in this thesis, without judgement, there can be an unnecessary fear of inefficiency or, alternatively, an unwarranted acceptance of being efficient at the wrong thing. As CEA is often performed with narrow measures of effectiveness, decision-makers should always judge whether or not the effectiveness measure in question offers 'appropriate' information to each particular decision-making process.
Sometimes a single efficiency measure will simply *not* be appropriate, even as only a 'starting point' for discussion. Because of the inherent variation in the *quality* of information provided by health-economic evaluations, generalisations about its usefulness as a concept, as well as the usefulness of any information it produces, should not be made. Instead, each CEA must be judged on its own merits in the context to which it is being applied.

7.1.2 Simplification and meaning

Assessment is fundamental to planning, but the extent to which 'measurement' contributes to planning should depend on the quality and relevance of that measurement. By restricting the scope of CEA, consistency between evaluations using CEA can be increased. However, this increase in consistency will often come at the expense of CEA no longer faithfully addressing policy questions. This should also be kept in mind when interpreting indicators such as those produced via CEA. On the other hand, classical health-economic evaluation (CHEE) is typically more likely to require serious thought and discussion, but may be easier to defend to stakeholders due to its flexibility in what dimensions of value it can consider. Extra-welfarist health-economic evaluation (EWHEE) may place less of a cognitive burden on decision-makers. but may be harder to defend, due to its often misleading one-size-fits-all approaches.

7.1.3 Known unknowns about economic evaluation?

Information stemming from health-economic evaluation should not be allowed to shut down or excessively crowd out awareness and debate about other important considerations. Nor should health-economic evaluation be allowed to crowd out a wider discussion of what society might want from its health and social care systems, or how much they may want to pay for these in total. For example, there is likely to be a balance to be struck between providing care to patients 'at any price' and fuelling the profits of national or multi-national private-sector companies. There are numerous impacts which may be considered to be legitimate or relevant, but some, like shareholder dividends from the medical industry, seem conspicuous by their absence from the majority of recent health-economic evaluation literature. Other, often relatively hidden or ignored, impacts are those related to the environment, the potentially vested interests of healthcare professionals, and the vested interests of researchers, to name but three. Prioritisation systems might be wise to take into account sustainability in all its forms, both within and without the health and social care sectors.

This sub-section leads to the **first key recommendation:** In order to be put into the correct context, claims about CEA, as a truncated form of efficiency indicator, should always be accompanied by an assessment of what is covered by the efficiency indicator, how well it is measured and what is left out. The credibility and appropriateness of CEA have central a role to play in defining its ability to contribute both to decision-making processes and to more comprehensive valuation. Only when CEA's limitations are clear can broader, higher-level efficiency, i.e., value, be properly assessed.

7.2 Evidence about prostate-cancer screening?

On the basis of the findings added by this study, evidence is still inadequate, plenty is still not known. Ambiguity remains, even after 20 years of the FinRSPC, as the analysis undertaken here is, given the uncertain public-health impact of prostate-cancer screening, still a relatively small sample population. Due to the relatively short follow-up (in part given the nature of prostate cancer as a disease), and given the probable influence of context, i.e., of socio-economic, environmental and clinical variation, impacts remain

uncertain. That said, the CEA of the FinRSPC so far finds there *may* be causes for concern about the impact of prostate-cancer screening on *some* diagnosed men. However, here the claim will be made that the direction, size and nature of many of the impacts of prostate-cancer screening are not fully able to be described with sufficient certainty, at least not at this point in time and not from this single RCT alone.

Evidence, even in the presence of many unknowns can be informative, but researchers should make it clearer that the evidence on prostate-cancer screening is still deficient in many ways and it is unlikely that stakeholders can be 'fully' informed. It should be clear that 'evidence-based' or 'evidence-informed' does not necessarily mean 'well informed', because it seems unlikely to be possible to fully understand the risks and benefits of screening at this point in time. Rather than unthinking repetition (or synthesis) of earlier methods or studies, an investment of time and resources in RCT and registry-based CEAs could check if the focus of earlier studies has been misplaced.

7.2.1 A need for studies in future?

Given the uncertainties revealed by this study, further studies trying to assess the cost-effectiveness of prostate-cancer screening should be wary about their analyses until more is known about the extent to which the findings from the current study, especially in terms of men likely overdiagnosed in the screening arm, might be generalisable. If registers do not contain information on HRQoL or other health-related outcomes, or if all treatments are not fully and consistently recorded, then there may be large gaps in knowledge within any future research. Robust evaluation would require the correct information to be collected at regular and appropriate intervals, from a sufficiently large and representative target group.

Given the above, the **second key recommendation** is**:** In order to be better informed about the possible benefits and harms of prostate-cancer screening, analysis would first need be undertaken to investigate if robust information about overdiagnosis or all-cause mortality could be produced. If the possibilities for the production of such information are slim, the chronic lack of robust information about many of the costs and benefits of screening should be more widely acknowledged.

7.3 CEA evidence and decisions: what is missing?

Given the difficulties with application of CEA to as complex a question as prostate-cancer mass screening, we should not be surprised by the lack of adequate evidence, just humble about what we know, and honest about what we don't know.

It should be noted that the information that comes from health-economic and epidemiological studies has a tendency to be bound to the conventions of that type of research. If cancer screening is evaluated in epidemiology primarily through analysis of target-cancer-specific mortality, it may miss other important effects. If health-economic evaluation follows such conventions or, say, the conventions of extra-welfarist modelling, and if any of those conventions are misleading in some way, then errors may remain undiscovered.

It also seems useful here to spell out the danger of deciding too hastily about prostate-cancer screening, and of appearing too certain about what is known. If the average effects, i.e. the indicators, measures and estimates, which often underpin health-economic evaluation are deficient or otherwise misleading, then the results of CEAs which use them may well be deficient or otherwise misleading, too. Perhaps only a mix of disciplines, challenging and supporting each other to make progress by challenging research conventions can provide a more complete, nuanced and realistic analysis.

On some occasions formal analysis will undoubtedly play a major role in decision-making, on some occasions only a minor role: this uncertainty about the role of CEA, the variability in its likely importance, is itself an important message. This message is repeated here in an attempt to reduce the overselling and arrogance which is problematic within the field of health-economic evaluation and beyond.

Sometimes CEA could be a starting point for discussions, sometimes its results could even be the deciding factor in discussions, sometimes its results mislead, but on all occasions its qualities should be checked. For this reason the **third key recommendation** is that CEA, classical health-economic evaluation, as well as deliberative analysis, would all be used as part of planning processes.

7.3.1 New interpretations of information from CEA?

The thesis presented here is that, firstly, interpretation of any CEA requires an understanding about which of the different approaches to economic evaluation is being used and what that particular method can offer *in principle*. Secondly, it requires an understanding of the extent the chosen method or methods have succeeded in producing useful information on a case-by-case basis *in practice*. Only then can that information gained from those assessment processes be appraised in light of other considerations. Each application of economic evaluation should be subjected to thorough interpretation by each stakeholder, being supplemented by other information when necessary. Given the complexities of many of the topics to which economic evaluation is applied in the field of health, simple interpretation will rarely be possible.

Stakeholders should be aware of the variability in the quality of economic evaluations, as well as the absence of any quality-assessment tool which would be capable of distinguishing between economic evaluations that are highly relevant, and those that are highly misleading. Given the apparent lack of suitable quality-assessment *tools*, stakeholders could benefit from processes which include deliberative analysis, which, in turn, relies on *judgment*.

References

Adami, Kalager, Valdimarsdottir, Bretthauer & Ioannidis (2018). Time to abandon early detection cancer screening. *European Journal of Clinical Investigation*, e13062.

Afzali & Karnon (2015). Exploring structural uncertainty in model-based economic evaluations. *Pharmacoeconomics*, 33, 435-443.

Alexander (2008). *The mantra of efficiency: from waterwheel to social control*. Baltimore: The Johns Hopkins University Press.

Alexander (2009). The Concept of Efficiency: An Historical Analysis. (In Meijers (Ed.), *PhilosophIny of Technology and Engineering Sciences* (pp. 1007-1030). Amsterdam: North-Holland).

Anderson (2010). Systematic reviews of economic evaluations: utility or futility? *Health Economics*, 19, 350-364.

Andriole, Crawford, Grubb, Buys, Chia, Church, et al. (2009). Mortality results from a randomized prostate-cancer screening trial. *The New England Journal of Medicine*, 360, 1310-1319.

Angell (1986). Publish or Perish: A Proposal. *Annals of Internal Medicine*, 104, 261-2.

Arnsrud Godtman, Holmberg, Lilja, Stranne & Hugosson (2015). Opportunistic testing versus organized prostate-specific antigen screening: outcome after 18 years in the Goteborg randomized population-based prostate cancer screening trial. *European Urology*, 68, 354-360.

Arrow (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review*, 53, 941-973.

Arrow & Lind (1970). Uncertainty and the Evaluation of Public Investment Decisions. *The American Economic Review*, 60, 364-378.

Ashmore, Mulkay & Pinch (1989). *Health and efficiency: A sociology of health economics*. Milton Keynes: Open University Press.

Auvinen & Hakama. (2017). Cancer Screening: Theory and Applications. (In Quah (Ed.), *International Encyclopedia of Public Health* (pp. 389-405): Elsevier).

Auvinen & Hugosson (2003). The rationale for the ERSPC trial: will it improve the knowledge base on prostate cancer screening? *BJU International*, 92, 14-16.

Auvinen, Moss, Tammela, Taari, Roobol, Schröder, et al. (2016). Absolute Effect of Prostate Cancer Screening: Balance of Benefits and Harms by Center within the European Randomized Study of Prostate Cancer Screening. *Clinical Cancer Research*, 22, 243-249.

Auvinen, Rannikko, Taari, Kujala, Mirtti, Kenttämies, et al. (2017). A randomized trial of early detection of clinically significant prostate cancer (ProScreen): study design and rationale. *European Journal of Epidemiology*, 32, 521-527.

Backhouse (2005). The Rise of Free Market Economics: Economists and the Role of the State since 1970. *History of Political Economy*, 37, 355-392.

Baker, Kramer & Prorok (2002). Statistical issues in randomized trials of cancer screening. *BMC Medical Research Methodology*, 2, 11.

Baltussen, Jansen, Bijlmakers, Grutters, Kluytmans, Reuzel, et al. (2017). Value Assessment Frameworks for HTA Agencies: The Organization of Evidence-Informed Deliberative Processes. *Value in Health*, 20, 256-260.

Banta (2018). Perspective: some conclusions from my life in health technology assessment. *International Journal of Technology Assessment in Health Care*, 34, 131-133.

Bauer (1992). *Scientific literacy and the myth of the scientific method*. Urbana: University of Illinois Press.

Bell, Del Mar, Wright, Dickinson & Glasziou (2015). Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *International Journal of Cancer*, 137, 1749-1757.

Bentham (1843). A Manual of Political Economy. (Written 1790–95, published under the Superintendence of his Executor as: *The works of Jeremy Bentham*. Volume 3. Edinburgh: William Tait).

Bergman & Litwin (2014). The Henderson-Hasselbalch Equation for Urologists. *European Urology*, 65, 48-49.

Berry (2017). *Morality and Power: On Ethics, Economics and Public Policy*. Cheltenham: Edward Elgar Publishing Limited.

Biderman (1966). Social Indicators and Goals. (In Bauer (Ed.), *Social indicators*. (pp. 68-153) Cambridge, Mass.: M.I.T. Press).

Birch (1997). As a matter of fact: evidence-based decision-making unplugged. *Health Economics*, 6, 547-559.

Birch (2002). Making the problem fit the solution: Evidence-based decision making and "Dolly" economics. (In Donaldson, Mugford & Vale (Eds.), *Evidence-based Health Economics: from Effectiveness to Efficiency in Systematic Review* (pp. 133-147). London: BMJ Books).

Birch & Donaldson (1987). Applications of cost-benefit analysis to health care: Departures from welfare economic theory. *Journal of Health Economics*, 6, 211-25.

Birch & Gafni (1996). Cost effectiveness and cost utility analyses: methods for the non-economic evaluation of health care programs and how we can do better. (In Geisler & Heller (Eds.), *Managing technology in health care* (pp. 51–67). New York: Wiley).

Birch & Gafni (2003). Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results. *Health Policy*, 64, 207-219.

Birch & Gafni (2004). Evidence-based health economics: Answers in search of questions? (In Kristiansen & Mooney (Eds.), *Evidence-based medicine: In its place* (pp. 50-61). London: Routledge).

Birch & Gafni (2006). The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. *Journal of Health Services Research and Policy*, 11, 46-51.

Birch & Gafni (2011). The Inconvenient Economic Truth: Benefits Forgone as an Input to Economic Evaluation and Implications for Decision-making. (In Rosen, Israeli & Shortell (Eds.), *Improving Health and Healthcare. Who is Responsible? Who is Accountable?* (pp. 601-622). Jerusalem, Israel: The Israel National Institute for Health Policy Research).

Black, Haggstrom & Welch (2002). All-Cause Mortality in Randomized Trials of Cancer Screening. *JNCI: Journal of the National Cancer Institute*, 94, 167-173.

Blaug (2007). The Fundamental Theorems of Modern Welfare Economics, Historically Contemplated. *History of Political Economy*, 39, 185-207.

Booth (2016). Might comparing trial-based and model-based estimates of costs be able to help when assessing the 'cost-effectiveness' of prostate-cancer screening? Health Economists' Study Group (HESG), Manchester, January 2016.

Booth (2019a). On value frameworks and opportunity costs in health technology assessment. *International Journal of Technology Assessment in Health Care*, 35, 367-372.

Booth (2019b). VP08: Can Health-Economic Evaluation Provide a Representation of 'Value For Money' For HTA? *International Journal of Technology Assessment in Health Care*, 35, 77.

Booth, Aronen & Mäkelä. (2017). Report on using cost-effectiveness as one of the criteria for defining health-care service choices [in Finnish]. Reports and Memorandums of the Ministry of Social Affairs and Health 2017:30. Helsinki: Ministry of Social Affairs and Health.

Booth, Rissanen, Tammela, Kujala, Stenman, Taari, et al. (2019). Register-based cost-effectiveness analysis of organised prostate-specific antigen screening for prostate cancer: Evidence from a randomised controlled trial. *PLOS ONE*, 14, e0224479.

Booth, Rissanen, Tammela, Määttänen, Taari & Auvinen (2014a). Health-related quality of life in the Finnish Trial of Screening for Prostate Cancer. *European Urology*, 65, 39-47.

Booth, Rissanen, Tammela, Määttänen, Taari & Auvinen (2014b). Reply from Authors re: Jonathan Bergman, Mark S. Litwin. The Henderson-Hasselbalch Equation for Urologists. Eur Urol 2014;65:48–9: Health-related Quality-of-life Scores from the Finnish Trial of Screening for Prostate Cancer Offer One Useful Contribution. *European Urology*, 65, 50-51.

Boulding (1966). The Ethics of Rational Decision. *Management Science*, 12, B161-9.

Brazier, Rowen, Lloyd & Karimi (2019). Future Directions in Valuing Benefits for Estimating QALYs: Is Time Up for the EQ-5D? *Value in Health*, 22, 62-68.

Briggs, Weinstein, Fenwick, Karnon, Sculpher, Paltiel, et al. (2012). Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Medical Decision Making*, 32, 722-732.

Brock, Daniels, Neumann & Siegel (2016). Ethical and Distributive Considerations. (In Neumann, Ganiats, Russell, Sanders, Siegel & Ganiats (Eds.), *Cost-effectiveness in health and medicine (second edition)* (pp. 319-342). Oxford: Oxford University Press).

Brockis, Marsden, Cole & Devlin (2016). A Review of NICE Methods Across Health Technology Assessment Programmes: Differences, Justifications and Implications. London: The Office of Health Economics (OHE).

Brommels & Sintonen (2001). Be generic and specific: quality of life measurement in clinical studies. *Annals of Medicine*, 33, 319-322.

Brousselle & Lessard (2011). Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. *Social Science & Medicine*, 72, 832-839.

Brouwer, van Baal, van Exel & Versteegh (2019). When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *The European Journal of Health Economics*, 20, 175-180.

Browne (1658). *Pseudodoxia Epidemica: Or, Enquiries Into Very Many Received Tenents, and Commonly Presumed Truths*. London.

Butterfield (1968). *Priorities in medicine*. Aylesbury, Bucks.: Nuffield Provincial Hospitals Trust.

Campbell & Pryce (2003). *Brick: a world history*. London: Thames and Hudson.

Cantillon (1755). *Essai sur la Nature du Commerce en General*. London: Fletcher, Gyles.

Caro (2009). Pursuing Efficiency: A Dead End for HTA? *Value in Health*, 12, S49.

Caro, Nord, Siebert, McGuire, McGregor, Henry, et al. (2010). IQWiG methods – a response to two critiques. *Health Economics*, 19, 1137-1138.

Carter, Degeling, Doust & Barratt (2016). A definition and ethical evaluation of overdiagnosis. *Journal of Medical Ethics*, 42, 705-714.

Carter, & Barratt (2017). What is overdiagnosis and why should we take it seriously in cancer screening? *Public Health Research & Practice,* 27, 3, e2731722

Cartwright & Hardie (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford Oxford University Press.

Churchman, Ackoff & Smith (1954). An Approximate Measure of Value. *Journal of the Operations Research Society of America*, 2, 172-187.

Ciatto, Zappa, Villers, Paez, Otto & Auvinen (2003). Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. *BJU International*, 92 Suppl 2, 97-100.

Clarke (2006). Evidence-Based Evaluation in Different Professional Domains: Similarities, Differences and Challenges. (In Shaw, Greene & Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 559-581). SAGE Publications Ltd).

Claxton (1999). The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, 18, 341-364.

Claxton, Briggs, Buxton, Culyer, McCabe, Walker, et al. (2008). Value based pricing for NHS drugs: an opportunity not to be missed? *BMJ*, 336, 251-254.

Coast (2004). Is economic evaluation in touch with society's health values? *BMJ*, 329, 1233-1236.

Coast (2017). A history that goes hand in hand: Reflections on the development of health economics and the role played by Social Science & Medicine, 1967–2017. *Social Science & Medicine*, 196,227-232.

Cochrane (1972 (2004)). *Effectiveness and efficiency: random reflections on health services*. London: Royal Society of Medicine Press.

Cochrane & Holland (1971). Validation of screening procedures. *British Medical Bulletin*, 27, 3-8.

Colvin (1987). *The Economic Ideal in British Government: Calculating Costs and Benefits in the 1970s*. Manchester: Manchester University Press.

Cookson & Claxton (Eds.) (2012). *The humble economist: Tony Culyer on health, health care and social decision making*. York: Office of Health Economics and University of York.

Cookson, Goddard, Sheldon & Maynard (2016). *Maynard matters: critical thinking on health policy*. York: University of York.

Crile (1955). A Plea against blind fear of cancer: An experienced surgeon says that excessive worry leads to costly tests, undue suffering. Life, 39, 128-142.

Culyer (1976). *Need and the National Health Service: Economics and social choice*. Bath: Martin Robertson.

Culyer (1981). Health, economics and health economics. (In Van der Gaag & Perlman (Eds.), *Health, economics, and health economics* (pp. 3-11). Oxford: North Holland Publishing Company).

Culyer (1989). The Normative Economics of Health Care Finance and Provision. *Oxford Review of Economic Policy*, 5, 34-58.

Culyer (2008). Resource allocation in health care: Alan Williams' decision maker, the authority and Pareto. (In Mason & Towse (Eds.), *The Ideas and Influence of Alan Williams: Be Reasonable - Do it My Way!* (pp. 57-74). Chichester: John Wiley & Sons).

Culyer (2014). *The Dictionary of Health Economics (third edition)*. London: Edward Elgar.

Culyer (2015). Four Issues in Cost-Effectiveness Analysis and Health Technology Assessment: a View from the Touch-line. (In del Llano-Señarís & Campillo-Artero (Eds.), *Health Technology Assessment and Health Policy Today: A Multifaceted View of their Unstable Crossroads* (pp. 77-94): Springer International Publishing).

Culyer (2016a). Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. *Health Economics, Policy and Law*, 11, 415-432.

Culyer (2016b). Cost-effectiveness thresholds: a comment on the commentaries. *Health Economics, Policy and Law*, 11, 445-447.

Culyer (2018). Cost, context, and decisions in health economics and health technology assessment. *International Journal of Technology Assessment in Health Care*, 34, 434-441.

Culyer, McCabe, Briggs, Claxton, Buxton, Akehurst, et al. (2007). Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *Journal of Health Services Research and Policy*, 12, 56-58.

Culyer, Williams & Lavers (1971). Social indicators: health. *Social Trends*, 2, 31-42.

Daddario (1967). *House of Representatives Bill 6698*. U.S. House of Representatives, Ninetieth Congress, first session. Washington: U.S. Govt. Print. Off.

Daddario (1968). Technology Assessment -- A Legislative View. *The George Washington Law Review*, 36, 1044-1059.

Daniels & van der Wilt (2016). Health technology assessment, deliberative process, and ethically contested issues. *International Journal of Technology Assessment in Health Care*, 32, 10-15.

Dantzig (1951). *Activity analysis of production and allocation: proceedings of a conference*. New York: Wiley.

de Mayerne Turquet (1611). *La Monarchie aristodémocratique, ou le Gouvernement composé et meslé des trois formes de légitimes républiques*. Paris: J. Berjon.

de Santos (2009). Fact-Totems and the Statistical Imagination: The Public Life of a Statistic in Argentina 2001. *Sociological Theory*, 27, 466-489.

Deaton & Cartwright (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21.

Dempsey (2019). *Anticipation and Medicine: A Critical Analysis of the Science, Praxis and Perversion of Evidence Based Healthcare*. London: Routledge.

Department of Health (1993). Research for health: HMSO.

Dewey (1929). *The quest for certainty: A study of the relation of knowledge and action*. London: George Allen & Unwin.

Dobrow, Goel & Upshur (2004). Evidence-based health policy: context and utilisation. *Social Science & Medicine*, 58, 207-217.

Drew (1967). HEW Grapples with PPBS. *The Public Interest*, 8, 9-29.

Dror (1970). Prolegomena to policy sciences. *Policy Sciences* (pp. 135-150): Springer Netherlands.

Drost, Rannikko, Valdagni, Pickles, Kakehi, Remmers, et al. (2018). Can active surveillance really reduce the harms of overdiagnosing prostate cancer? A reflection of real life clinical practice in the PRIAS study. *Translational andrology and urology*, 7, 98-105.

Drummond (1978). Evaluation and the National Health Service. (In Culyer & Wright (Eds.), *Economic aspects of health services* (pp. 67-83). London: Martin Robertson).

Drummond (1980). Chapter 6: Decision rules and decisions. (In Drummond, *Principles of Economic Appraisal in Health Care*. Oxford: Oxford University Press).

Drummond (1981). Welfare economics and cost benefit analysis in health care[1]. *Scottish Journal of Political Economy*, 28, 125-145.

Drummond & Jefferson (1996). Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ*, 313, 275-283.

Drummond & McGuire (Eds.) (2001). *Economic evaluation in health care: merging theory with practice*. Oxford: Oxford University Press.

Drummond, Sculpher, Claxton, Stoddart & Torrance (2015). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press.

Duffy, Nagtegaal, Wallis, Cafferty, Houssami, Warwick, et al. (2008). Correcting for Lead Time and Length Bias in Estimating the Effect of Screen Detection on Cancer Survival. *American Journal of Epidemiology*, 168, 98-104.

Eastwood & Maynard (1990). Treating Aids: Is it ethical to be efficient? (In Baldwin, Godfrey & Propper (Eds.), *Quality of Life: Perspectives and Policies* (pp. 231-249). London: Routledge).

Editorial (1899). Vaccination in India. *BMJ*, 1, 1341.

Esserman, Thompson, Reid, Nelson, Ransohoff, Welch, et al. (2014). Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncology*, 15, e234-242.

European Network for Health Technology Assessment (EUnetHTA) (2015). Methods for health economic evaluations—a guideline based on current practices in Europe. http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/2015-04-29-ECO-GL_Final%20version_0.pdf

Evans (1984). *Strained mercy: the economics of Canadian health care*. Toronto: Butterworths.

Evers, Goossens, de Vet, van Tulder & Ament (2005). Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *International Journal of Technology Assessment in Health Care*, 21, 240-245.

Fall, Fang, Mucci, Ye, Andren, Johansson, et al. (2009). Immediate risk for cardiovascular events and suicide following a prostate cancer diagnosis: prospective cohort study. *PLOS Medicine*, 6, e1000197.

Fang, Keating, Mucci, Adami, Stampfer, Valdimarsdóttir, et al. (2010). Immediate Risk of Suicide and Cardiovascular Death After a Prostate Cancer Diagnosis: Cohort Study in the United States. *JNCI: Journal of the National Cancer Institute*, 102, 307-314.

Feeny, Krahn, Prosser & Salomon (2016). Chapter 7. Valuing Health Outcomes. (In Neumann, Ganiats, Russell, Sanders, Siegel & Ganiats (Eds.), *Cost-effectiveness in health and medicine (second edition)* (pp. 167-199). Oxford: Oxford University Press).

Feldstein (1963). Economic Analysis, Operational Research, and the National Health Service. *Oxford Economic Papers*, 15, 19-31.

Finne, Fallah, Hakama, Ciatto, Hugosson, de Koning, et al. (2010). Lead-time in the European Randomised Study of Screening for Prostate Cancer. *European Journal of Cancer*, 46, 3102-3108.

Finne, Stenman, Määttänen, Mäkinen, Tammela, Martikainen, et al. (2003). The Finnish trial of prostate cancer screening: where are we now? *BJU International*, 92 Suppl. 2, 22-26.

Fisher (1965). The role of cost-utility analysis in program budgeting. (In Novick (Ed.), *Program budgeting; program analysis and federal budget* (pp. 61-78). Cambridge, Mass.: Harvard University Press).

Fox, Grimm & Rute (2016). *An Introduction to Evaluation*. Thousand Oaks, California: SAGE.

Franklin (1748). Advice to a Young Tradesman, written by an old one. (In Fisher (Ed.), *The American Instructor: or Young Man's Best Companion* (pp. 375–7). Philadelphia, PA: Benjamin Franklin & David Hall).

Franklin (1772). A letter to Joseph Priestley (Fellow of the Royal Society), London.

Franklin (1842). *Memoirs of Benjamin Franklin*. New York: Harper & Brothers.

Furlan, Malmivaara, Chou, Maher, Deyo, Schoene, et al. (2015). 2015 Updated Method Guideline for Systematic Reviews in the Cochrane Back and Neck Group. *Spine*, 40, 1660-1673.

Gafni & Birch (2006). Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Social Science & Medicine* (1982), 62, 2091-2100.

Galbraith (1973). How prices are set. (In Galbraith, *Economics and the public purpose* (pp. 106-117). Boston: Houghton Mifflin).

Galbraith (1979). *The New Industrial State*. New York: New American Library.

Garber & Phelps (1997). Economic foundations of cost-effectiveness analysis. *Journal of Health Economics*, 16, 1-31.

Garrison, Neumann, Willke, Basu, Danzon, Doshi, et al. (2018). A Health Economics Approach to US Value Assessment Frameworks—Summary and Recommendations of the ISPOR Special Task Force Report [7]. *Value in Health*, 21, 161-165.

Gerard (1992). Cost-utility in practice: A policy maker's guide to the state of the art. *Health Policy*, 21, 249-279.

Glick, Kinosian & Schulman (1994). Decision Analytic Modeling: Some Uses in the Evaluation of New Pharmaceuticals. *Therapeutic Innovation & Regulatory Science*, 28, 691-707.

Godin (2007). Science, accounting and statistics: The input–output framework. *Research Policy*, 36, 1388-1403.

Goldenberg (2006). On evidence and evidence-based medicine: Lessons from the philosophy of science. *Social Science & Medicine*, 62, 2621-2632.

Green (1894). Pain-Cost and Opportunity-Cost. *The Quarterly Journal of Economics*, 8, 218-229.

Green (1990). Health economics: are we being realistic about its value? *Health Policy and Planning*, 5, 274-279.

Green & Barker (1988). Priority setting and economic appraisal: whose priorities--the community or the economist? *Social Science & Medicine*, 26, 919-929.

Gusmano & Callahan (2011). "Value for Money": Use With Care. *Annals of Internal Medicine*, 154, 207-208.

Gyrd-Hansen (2014). Efficiency in Health Care, Concepts of. (In Culyer (Ed.), *Encyclopedia of health economics* (pp. 276-271). Burlington: Elsevier Science).

Gøtzsche (2019). *Death of a whistleblower and Cochrane's moral collapse*. People's Press.

Hakama (1991). Screening. (In *Oxford textbook of public health* (pp. 91–105). Oxford: Oxford University Press).

Hammond (1958). *Benefit-cost analysis and water-pollution control*. Stanford, CA: Food Research Institute, Stanford University.

Hammond (1991). Frank Knight's Antipositivism. *History of Political Economy*, 23, 359-381.

Hardie, Parker, Norman, Eeles, Horwich, Huddart, et al. (2005). Early outcomes of active surveillance for localized prostate cancer. *BJU International*, 95, 956-960.

Hausman, McPherson & Satz (2017). *Economic analysis, moral philosophy and public policy*. Cambridge: Cambridge University Press.

Haythorn & Ablin (2011). Prostate-specific antigen testing across the spectrum of prostate cancer. *Biomarkers in Medicine*, 5, 515-526.

Heijnsdijk, Csanádi, Gini, ten Haaf, Bendes, Anttila, et al. (2019). All-cause mortality versus cancer-specific mortality as outcome in cancer screening trials: A review and modeling study. *Cancer Medicine*, 8, 6127-6138.

Heijnsdijk, de Carvalho, Auvinen, Zappa, Nelen, Kwiatkowski, et al. (2015). Cost-effectiveness of prostate cancer screening: A simulation study based on ERSPC data. *Journal of the National Cancer Institute*, 107.

Heintz, Gerber-Grote, Ghabri, Hamers, Rupel, Slabe-Erker, et al. (2016). Is There a European View on Health Economic Evaluations? Results from a Synopsis of Methodological Guidelines Used in the EUnetHTA Partner Countries. *Pharmacoeconomics*, 34, 59-76.

Hicks (1986). A discipline not a science. (In Hicks (Ed.), *Collected essays on economic theory*. Vol. 3, Classics and moderns (pp. 365-375). Oxford: Blackwell).

Hitch (1953). Sub-Optimization in Operations Problems. *Journal of the Operations Research Society of America*, 1, 87-99.

Hitch (1967). *Decision-making for defense*. Berkeley (CA): University of California Press.

Hitch (1971). *On the choice of objectives in systems studies*. RAND paper series, P-1955. Santa Monica, CA.: RAND Corporation.

HM Treasury (2018). *The Green Book: central government on appraisal and evaluation*. London: HMSO.

Hodgson (2001). *How Economics Forgot History: The Problem of Specificity in Social Science*. Taylor & Francis Group.

Howick (2011). *The philosophy of evidence-based medicine*. Chichester, West Sussex, UK: Wiley-Blackwell, BMJ Books.

Hugosson, Carlsson, Aus, Bergdahl, Khatami, Lodding, et al. (2010). Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. *The Lancet Oncology*, 11, 725-732.

Hugosson, Roobol, Månsson, Tammela, Zappa, Nelen, et al. (2019). A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *European Urology*, 76, 43-51.

Hurley (1998). Chapter 16: Welfarism, extra-welfarism and evaluative economic analysis in the health sector. (In Barer, Getzen & Stoddart (Eds.), *Health, health care and health economics: perspectives on distribution* (pp. 373-395). Chichester: Wiley).

Husereau, Drummond, Petrou, Carswell, Moher, Greenberg, et al. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS) – explanation and elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value in Health*, 16, 231-250.

Häkkinen (2011). The PERFECT project: measuring performance of health care episodes. *Annals of Medicine*, 43, S1-S3.

Hämäläinen, Kovasin & Räikkonen (2019). Ehdotukset sote-ohjauksen mittareista ja tietopohjan varmentamisesta. Sosiaali- ja terveysministeriön raportteja ja muistioita 2019:33. Helsinki: Sosiaali- ja terveysministeriö.

Joint Economic Committee (1967). *The Planning-Programming-Budgeting System: Progress and Potentials*. Congress of the United States, 90th Congress, First Session.

Johansson, Steineck, Holmberg, Johansson, Nyberg, Ruutu, et al. (2011). Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. *Lancet Oncology,* 12, 891-899.

Jonsen (1986). Bentham in a box: technology assessment and health care allocation. Law, *Medicine and Health Care*, 14, 172-174.

Kalager, Wieszczy, Lansdorp-Vogelaar, Corley, Bretthauer & Kaminski (2018). Overdiagnosis in Colorectal Cancer Screening: Time to Acknowledge a Blind Spot. *Gastroenterology*, 155, 592-595.

Kaltenthaler, Tappenden & Paisley (2013). Reviewing the Evidence to Inform the Population of Cost-Effectiveness Models within Health Technology Assessments. *Value in Health*, 16, 830-836.

Karlsson & Johannesson (1996). The decision rules of cost-effectiveness analysis. *Pharmacoeconomics*, 9, 113-120.

Kay & King (2020). *Radical Uncertainty: Decision-making for an unknowable future*. The Bridge Street Press.

Kelleher (2014). Efficiency and Equity in Health: Philosophical Considerations. (In Culyer (Ed.), *Encyclopedia of health economics* (pp. 259-266). Burlington: Elsevier Science).

Keller, Gericke, Whitty, Yaxley, Kua, Coughlin, et al. (2017). A Cost-Utility Analysis of Prostate Cancer Screening in Australia. *Applied Health Economics and Health Policy*, 15, 95-111.

Keynes (1924). Alfred Marshall, 1842-1924. *The Economic Journal*, 34, 311-372.

Kilpeläinen, Pogodin-Hannolainen, Kemppainen, Talala, Raitanen, Taari, et al. (2017). Estimate of opportunistic prostate specific antigen testing in the Finnish Randomized Study of Screening for Prostate Cancer. *The Journal of Urology*, 198, 50-57.

Kilpeläinen, Tammela, Malila, Hakama, Santti, Määttänen, et al. (2013). Prostate cancer mortality in the Finnish Randomized Screening Trial. *Journal of the National Cancer Institute*, 105, 719-725.

Klarman (1965). *The Economics of Health*. New York: Columbia University Press.

Klarman (1967). Present status of cost-benefit analysis in the health field. American *Journal of Public Health and the Nations Health*, 57, 1948-1953.

Klarman (1982). The Road to Cost-Effectiveness Analysis. The Milbank Memorial Fund Quarterly. *Health and Society*, 60, 585-603.

Klein (1989). The role of health economics. *BMJ*, 299, 275-276.

Knight (1923). The Ethics of Competition. *The Quarterly Journal of Economics*, 37, 579-624.

Knight (1951). The Rôle of Principles in Economics and Politics. *The American Economic Review*, 41, 1-29.

Knight (1956). *On the history and method of economics: Selected essays*. Chicago: University of Chicago Press.

Kobelt (2013). Health economics : an introduction to economic evaluation. London: Office of Health Economics.

Konu, Rissanen, Ihantola & Sund (2009). "Effectiveness" in Finnish healthcare studies. *Scandinavian Journal of Public Health*, 37, 64-74.

Kothari & Birch (1998). Concepts of Rigour and Implications for Health Services Research. *Journal of Health Services Research and Policy*, 3, 121-123.

Krahn (1996). Cost-effectiveness analysis: are the outputs worth the inputs? *ACP Journal Club*, 124, 82.

Krahn, Naglie, Naimark, Redelmeier & Detsky (1997). Primer on medical decision analysis: Part 4 – Analyzing the model and interpreting the results. *Medical Decision Making*, 17, 142-151.

Kristiansen & Gyrd-Hansen (2006). Communicating treatment effectiveness in the context of chronic disease processes. *Expert Review of Pharmacoeconomics and Outcomes Research*, 6, 673-679.

Kristiansen & Mooney (2004). Evidence-based medicine: Method, collaboration, movement or crusade? (In Kristiansen & Mooney (Eds.), *Evidence-based medicine: In its place* (pp. 1-19). London: Routledge).

Kurtz & Snowden (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal*, 42, 462-483.

Laine (2014). Kustannusvaikuttavuus ei yksin riitä priorisoinnin perustaksi [in Finnish]. *Duodecim*, 130, 2094–2098.

Laing (Ed.) (1972). Evaluation in the Health Services. London: Office of Health Economics.

Lessard & Birch (2010). Complex problems or simple solutions? Enhancing evidence-based economics to reflect reality. (In Shemilt, Mugford, Vale, Marsh & Donaldson (Eds.), *Evidence-based Decisions and Economics: Health care, social welfare, education and criminal justice* (pp. 162-172). London: BMJ Books).

Liliemark, Lööf, Befrits, Back & Sandman (2016). [The willingness to pay for new drugs is based on ethical principles: The New Therapies Council interprets the parliamentary framework for prioritisation]. *Lakartidningen*, 113, D4WP.

Loeb, Bjurlin, Nicholson, Tammela, Penson, Carter, et al. (2014). Overdiagnosis and Overtreatment of Prostate Cancer. *European Urology*, 65, 1046-1055.

Loomes & McKenzie (1989). The use of QALYs in health care decision making. *Social Science & Medicine*, 28, 299-308.

Lord, Laking & Fischer (2004). Health care resource allocation: is the threshold rule good enough? *Journal of Health Services Research and Policy*, 9, 237-245.

Lowe (2008). Value for money and the valuation of public sector assets. London: HM Treasury.

Lübbe (2011). Dissenting opinion. (In German Ethics Council (Ed.), *Medical Benefits and Costs in Healthcare: The Normative Role of their Evaluation* (pp. 96-121). Berlin: Opinion).

Malm (1999). Medical screening and the value of early detection. When unwarranted faith leads to unethical recommendations. *Hastings Center Report*, 29, 26-37.

Manski (2019). The lure of incredible certitude. *Economics and Philosophy*, 1-30.

Manvar, Pruthi, Wallen & Nielsen (2013). Epidemiology of Prostate Cancer. (In Tewari (Ed.), *Prostate Cancer: A Comprehensive Perspective* (pp. 285-300). London: Springer-Verlag).

Marsh, Goetghebeur, Thokala & Baltussen (Eds.) (2017). *Multi-Criteria Decision Analysis to Support Healthcare Decisions*: Springer.

Mason & Towse (2008). *The ideas and influence of Alan Williams - Be reasonable - Do it my way!* Oxford: Radcliffe.

Maynard (1999). Rationing health care: an exploration. *Health Policy*, 49, 5-11.

Maynard & Bloor (1997). Regulating the pharmaceutical industry. *BMJ*, 315, 200-201.

Maynard & Sheldon (1997). Health Economics: has it Fulfilled its Potential? (In Maynard, Chalmers & Cochrane (Eds.), *Non-random reflections on health services research: on the 25th anniversary of Archie Cochrane's Effectiveness and efficiency* (pp. 149-165). London: BMJ Publishing Group).

Mazzucato (2018). *The Value of Everything: Making and Taking in the Global Economy*. London: Penguin (Allen Lane).

McCabe, Claxton & Culyer (2008). The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics*, 26, 733-744.

McDonald, Charles, Elit & Gafni (2016). Challenges in striving to simultaneously achieve multiple resource allocation goals: the pan-Canadian Oncology Drug Review (pCODR) example. *Journal of Market Access & Health Policy*, 4, 31463.

McGuire. (2001). Theoretical concepts in the economic evaluation of health care. (In Drummond & McGuire (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice* (pp. 22–45). Oxford: Open University Press).

McIntosh, Clarke, Frew & Louviere (Eds.) (2010). *Applied Methods of Cost-Benefit Analysis in Health Care*. Oxford: Oxford University Press.

McIntosh, Donaldson & Ryan (1999). Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. *Pharmacoeconomics*, 15, 357-367.

McKean & Moore (1972). Uncertainty and the Evaluation of Public Investment Decisions: Comment. *The American Economic Review*, 62, 165-167.

McPake, Normand & Smith (2013). *Health economics: an international perspective*. London; New York: Routledge Taylor & Francis Group.

Meade & Hitch (1938). *Introduction to economic analysis and policy*. New York: Oxford University Press.

Melnikow, LeFevre, Wilt & Moyer (2013). Counterpoint: Randomized Trials Provide the Strongest Evidence for Clinical Guidelines: The US Preventive Services Task Force and Prostate Cancer Screening. *Medical Care*, 51, 301-303.

Meltzer (1997). Accounting for future costs in medical cost-effectiveness analysis. *Journal of Health Economics*, 16, 33-64.

Miller (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review*, 63, 81-97.

Miller, Madalinska, Church, Crawford, Essink-Bot, Goel, et al. (2001). Health-related quality of life and cost-effectiveness studies in the European randomised study of screening for prostate cancer and the US Prostate, Lung, Colon and Ovary trial. *European Journal of Cancer*, 37, 2154-2160.

Ministry of Social Affairs and Health (Finland). (2003). Annex to the decree of the Ministry of Social Affairs and Health concerning the Health Insurance Act.

Mirowski (1990). Learning the Meaning of a Dollar: Conservation Principles and the Social Theory of Value in Economic Theory. *Social Research*, 57, 689-717.

Mirowski (1999). Cyborg Agonistes: Economics Meets Operations Research in Mid-Century. *Social Studies of Science*, 29, 685-718.

Mishan (1971). *Cost-Benefit Analysis: An Informal Introduction*. London: George Allen & Unwin.

Mishan (1972). Uncertainty and the Evaluation of Public Investment Decisions: Comment. *The American Economic Review*, 62, 161-164.

Mishan (1982a). *Introduction to Political Economy*. London: Hutchinson.

Mishan (1982b). The new controversy about the rationale of economic evaluation. *Journal of Economic Issues*, 16, 29-47.

Mishan (1988). *Cost-Benefit Analysis: An Informal Introduction*. London: Allen & Unwin.

Mooney (1989). QALYs: are they enough? A health economist's perspective. *Journal of Medical Ethics*, 15, 148-152.

Mooney & Lange (1993). Ante-natal screening: What constitutes 'benefit'? *Social Science & Medicine,* 37, 873-878.

Moore, Baillie, Coppel & Royce. (2010). Supporting implementation of public health guidance: NICE experience. (In Killoran & Kelly (Eds.), *Evidence-based public health: effectiveness and efficiency* (pp. 411-424). Oxford: Oxford University Press).

Mugford (1989). The role of health economics. *BMJ*, 299, 741-741.

Mulkay, Pinch & Ashmore (1987). Colonizing the Mind: Dilemmas in the Application of Social Science. *Social Studies of Science*, 17, 231-256.

Muller (2018). *The tyranny of metrics*. Princeton: Princeton University Press.

Musgrove (1999). Public spending on health care: how are different criteria related? *Health Policy*, 47, 207-223.

Mäkinen, Karhunen, Aro, Lahtela, Määttänen & Auvinen (2008). Assessment of causes of death in a prostate cancer screening trial. *International Journal of Cancer*, 122, 413-417.

National Institute for Health and Care Excellence (2013). Guide to the methods of technology appraisal 2013. London: NICE.
(Accessed 19/03/20 via: https://www.nice.org.uk/process/pmg9/)

National Institute for Health and Care Excellence (2017). Our programmes. (Accessed 19/03/20: https://www.nice.org.uk/about/what-we-do/our-programmes)

National Institute for Health and Care Excellence. (2019). Guide to the processes of technology appraisal. (Accessed 19/03/20 via: https://www.nice.org.uk/process/pmg19/)

Naylor (1996). Cost-effectiveness analysis: are the outputs worth the inputs? *ACP Journal Club*, 124, A12-A14.

Neumann, Ganiats, Russell, Sanders, Siegel & Ganiats (Eds.) (2016). *Cost-effectiveness in health and medicine (second edition)*. Oxford: Oxford University Press.

Neumann, Kim, Trikalinos, Sculpher, Salomon, Prosser, et al. (2018). Future Directions for Cost-effectiveness Analyses in Health and Medicine. *Medical Decision Making*, 38, 767-777.

Neumann & Sanders (2017). Cost-Effectiveness Analysis 2.0. *New England Journal of Medicine*, 376, 203-205.

Norwegian Ministry of Health and Care Services. (2017). Principles for priority setting in health care — Summary of a white paper on priority setting in the Norwegian health care sector: Meld. St. 34 (2015–2016). Ministry of Health and Care Services.

O'Brien (1996). Economic evaluation of pharmaceuticals. Frankenstein's monster or vampire of trials? *Medical Care*, 34, DS99-DS108.

Office of Health Economics (Ed.) (1979). *Scarce Resources in Health Care*. London: Office of Health Economics.

Office of Technology Assessment (1976). *Development of medical technology: Opportunities for assessment*. Washington: U.S. Congress, Office of Technology Assessment : for sale by the Supt. of Docs., U.S. Govt. Print. Off.

Office of Technology Assessment (1980). *The Implications of Cost-Effectiveness Analysis of Medical Technology: Methodological Issues and Literature Review*. Washington, DC: U.S. Government Printing Office.

Orr & Wolff (2015). Reconciling cost-effectiveness with the rule of rescue: the institutional division of moral labour. *Theory and Decision*, 78, 525-538.

Oxford English Dictionary (2016). "*measure, n.*": Oxford University Press.

Paddison (1986). QALY's sale room (satirical cartoon). *Radical Community Medicine*, 24, 35.

Pataky, Gulati, Etzioni, Black, Chi, Coldman, et al. (2014). Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada. *International Journal of Cancer*, 135, 939-947.

Pawson (2006). *Evidence-Based Policy: A Realist Perspective*. London: Sage Publications.

Pawson, Greenhalgh, Harvey & Walshe (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research and Policy*, 10, 21-34.

Pigou (1947). *A study in public finance*. London: Macmillan and Co. Ltd.

Pole (1968). Economic Aspects of Screening for Disease. (In McKeown (Ed.), *Screening in medical care: Reviewing the evidence: A collection of essays with a preface by Lord Cohen of Birkenhead* (pp. 141-158). London: Oxford Univiversity Press).

Pole (1972). Mass Radiography: A Cost/Benefit Approach. (In Harberger, Haveman, Margolis, Niskanen, Turvey & Zeckhauser (Eds.), *Benefit-Cost Analysis 1971: An Aldine annual* (pp. 161-168). Chicago, Ill: Aldine).

Pole (1974). Programmes, priorities, and budgets. *British Journal of Preventive and Social Medicine*, 28, 191-195.

Porta (2014). *A Dictionary of Epidemiology*. Oxford University Press, USA.

Porter (1992). Quantification and the Accounting Ideal in Science. *Social Studies of Science*, 22, 633-651.

Porter (1995). *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton, N.J.: Princeton University Press.

Prest & Turvey (1965). Cost-Benefit Analysis: A Survey. *The Economic Journal*, 75, 683-735.

Priestley (1831). *Historical Account of the Navigable Rivers, Canals, and Railways of Great Britain: As a Reference to Nichols, Priestley & Walker's New Map of Inland Navigation*: Longman, Rees, Orme, Brown & Green.

Quade (1967). Introduction and overview. (In Goldman (Ed.), *Cost-effectiveness analysis: new approaches in decision-making* (pp. 1-16). New York: Praeger).

Quade (1970). On the limitations of Quantitative analysis. RAND paper series, P-4530. Santa Monica, CA.: RAND Corporation.

Quade (1971a). Chapter 14. Systems Analysis Techniques for Planning—Programming—Budgeting. (In Lyden & Miller (Eds.), *Planning Programming Budgeting: a systems approach to management* (pp. 292-312). Chicago: Markham Publishing Company).

Quade (1971b). A history of cost-effectiveness. RAND paper series, P-4557. Santa Monica, CA.: RAND Corporation.

Quade (1989). *Analysis for Public Decisions*. North-Holland: Elsevier.

Ramsey, Willke, Glick, Reed, Augustovski, Jonsson, et al. (2015). Cost-effectiveness analysis alongside clinical trials II – an ISPOR Good Research Practices Task Force report. *Value in Health*, 18, 161-172.

Ratcliffe & Gonzalez-Del-Valle (1988). Rigor in Health-Related Research: Toward an Expanded Conceptualization. *International Journal of Health Services*, 18, 361-392.

Relman (1980). The new medical-industrial complex. *The New England Journal of Medicine*, 303, 963-970.

Richardson, Iezzi, Khan, Chen & Maxwell (2016). Measuring the Sensitivity and Construct Validity of 6 Utility Instruments in 7 Disease Areas. *Medical Decision Making*, 36, 147-159.

Rickover (1967). Cost-effectiveness studies. (In Subcommittee on National Security and International Operations (Ed.), *Planning-programming-budgeting; Inquiry of the Subcommittee on National Security and International Operations* (pp. 599-608). Washington: U.S. Govt. Print. Off).

Roberts, Russell, Paltiel, Chambers, McEwan, Krahn, et al. (2012). Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Medical Decision Making*, 32, 678-689.

Robinson (1999). Limits to rationality: Economics, economists and priority setting. *Health Policy*, 49, 13-26.

Roobol (2015). Unorganized prostate-specific antigen-based screening for prostate cancer: more harm than benefit. When will we finally start to implement guidelines and risk assessment tools in clinical practice? *European Urology*, 68, 363-364.

Roth, Gulati, Gore, Cooperberg & Etzioni (2016). Economic analysis of prostate-specific antigen screening and selective treatment strategies. *JAMA Oncology*, 2, 890-898.

Rumford (1876). *The complete works of Count Rumford. Volume V. Of the fundamental principles on which general establishments for the relief of the poor may be formed in all countries*. London: Macmillan & Co.

Ruskin (1872). *Munera pulveris*. London: Smith, Elder & Co.

Russell & Sinha (2016). Strengthening Cost-Effectiveness Analysis for Public Health Policy. *American Journal of Preventive Medicine*, 50, S6-S12.

Sandel (2013). Market Reasoning as Moral Reasoning: Why Economists Should Re-engage with Political Philosophy. *Journal of Economic Perspectives*, 27, 121-140.

Schick (1966). The road to PPB: the stages of budget reform. *Public Administration Review*, 26, 243-258.

Schroder, Hugosson, Roobol, Tammela, Ciatto, Nelen, et al. (2009). Screening and prostate-cancer mortality in a randomized European study. The *New England Journal of Medicine*, 360, 1320-1328.

Schultze (1968). *The politics and economics of public spending*. Washington, D.C.: The Brookings Institution.

Schum (2009). A science of evidence: contributions from law and probability. *Law, Probability and Risk*, 8, 197-231.

Sculpher, Claxton & Akehurst (2005). It's just evaluation for decision-making: Recent developments in, and challenges for, cost-effectiveness research. (In Smith, Ginnelly & Sculpher (Eds.), *Health Policy and Economics: Opportunities and Challenges* (pp. 8-41). Maidenhead: Open University Press).

Sculpher, Claxton, Drummond & McCabe (2006). Whither trial-based economic evaluation for health care decision making? *Health Economics*, 15, 677-687.

Self (1970). "Nonsense on stilts": Cost-Benefit Analysis and the Roskill commission. *The Political Quarterly*, 41, 249-260.

Sheldon (1996). Problems of using modelling in the economic evaluation of health care. *Health Economics*, 5, 1-11.

Sheldon (2005). Discussion of "It's just evaluation for decision-making: Recent developments in, and challenges for, cost-effectiveness research". (In Smith, Ginnelly & Sculpher (Eds.), *Health Policy and Economics: Opportunities and Challenges* (pp. 33-36). Maidenhead: Open University Press).

Shemilt, McDaid, Marsh, Henderson, Bertranou, Mallander, et al. (2013). Issues in the incorporation of economic perspectives and evidence into Cochrane reviews. *Systematic Reviews*, 2:83.

Shemilt, Mugford, Byford, Drummond, Eisenstein, Knapp, et al. (2008). Incorporating Economics Evidence. *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 449-479): John Wiley & Sons, Ltd.

Shiell, Hawe & Gold (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ*, 336, 1281-1283.

Shin, Kim, Hwang, Lee, Lee & Ahn (2014). Economic evaluation of prostate cancer screening test as a national cancer screening program in South Korea. *Asian Pac J Cancer Prev.*, 15, 3383-3389.

Shteynshlyuger & Andriole (2011). Cost-Effectiveness of Prostate Specific Antigen Screening in the United States: Extrapolating From the European Study of Screening for Prostate Cancer. *The Journal of Urology*, 185, 828-832.

Silverberg & Ray (2018). Variations in Instructed vs. Published Word Counts in Top Five Medical Journals. *Journal of General Internal Medicine*, 33, 16-18.

Small & Mannion (2005). A hermeneutic science: health economics and Habermas. *Journal of Health Organization and Management*, 19, 219-235.

Smith (1761). *The Theory of Moral Sentiments*. (Second ed.): printed by A. Millar, London.

Smith (1776). *An inquiry into the nature and causes of the wealth of nations*. London: Strahan and Cadell.

Smith Jr, Walters, Brooks & Blackwell (1953). The theory of value and the science of decision. A summary. *Journal of the Operations Research Society of America*, 1, 103-113.

Specht (1960). RAND—A Personal View of Its History. *Operations Research*, 8, 825-839.

Stamey, Yang, Hay, McNeal, Freiha & Redwine (1987). Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *The New England Journal of Medicine*, 317, 909-916.

Steineck, Akre & Bill-Axelson (2019). Solid Science for the Upside but Lack of Solid Science for the Downside—Towards Cutting-edge Prostate-cancer Screening. *European Urology*, 76, 52-53.

Steiner (1965). Program budgeting: Business Contribution to Government Management. *Business Horizons*, 8, 43.

Stern (2004). Philosophies and types of evaluation research. (In Descy & Tessaring (Eds.), *The foundations of evaluation and impact research: Third report on vocational training research in Europe: background report*. Luxembourg: Office for Official Publications of the European Communities).

Steuart (1769). *Considerations on the interest of the county of Lanark in Scotland*. Glasgow: Printed for the author, by Robert Duncan.

Stevens (1959). Measurement, psychophysics, and utility. (In Churchman & Ratoosh (Eds.), *Measurement: definitions and theories* (pp. 18-63). New York: Wiley).

Strech & Tilburt (2008). Value judgments in the analysis and synthesis of evidence. *Journal of Clinical Epidemiology*, 61, 521-524.

Sugden & Williams (1978). *The Principles of Practical Cost-Benefit Analysis*. Oxford: Oxford University Press.

Tappenden (2014). Problem Structuring for Health Economic Model Development. (In Culyer (Ed.), *Encyclopedia of health economics* (pp. 168-179). Burlington: Elsevier Science).

The Danish Council of Ethics. (2001). Screening - a report. The Danish Council of Ethics 1999 Annual Report, Copenhagen.

The European network for Health Technology Assessment (EUnetHTA) project. (2016). HTA Core Model: version 3.0.

The Professional Society for Health Economics and Outcomes Research (ISPOR). Good Practice for Outcomes Research Reports (2003-2019). (Accessed 19/03/20: https://www.ispor.org/heor-resources/good-practices-for-outcomes-research)

Thokala, Ochalek, Leech & Tong (2018). Cost-Effectiveness Thresholds: the Past, the Present and the Future. *Pharmacoeconomics*, 36, 509-522.

Torrance (1970). A generalized cost-effectiveness model for the evaluation of health programs. Research report 101. Health Services Research. Hamilton, Ontario: McMaster University, Faculty of Business.

Trefethen (1954). A History of Operations Research. (In McCloskey & Trefethen (Eds.), *Operations Research for Management*, Vol. I.). Baltimore: Johns Hopkins Press.

Tsuchiya & Williams (2001). Welfare economics and economic evaluation. (In Drummond & McGuire (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice* (pp. 22–45). Oxford: Open University Press).

Tugwell, Bennett, Feeny, Guyatt & Haynes (1986). Chapter 4. A Framework for the Evaluation of Technology: The Technology Assessment Iterative Loop. (In Feeny, Guyatt & Tugwell (Eds.), *Health Care Technology: Effectiveness, Efficiency, and Public Policy* (pp. 41-56). Montreal, Canada: Institute for Research on Public Policy).

Tunis (2007). Reflections on science, judgment, and value in evidence-based decision making: A conversation with David Eddy. *Health Affairs*, 26, w500-w515.

US Preventive Services Task Force (2018). Screening for prostate cancer: US preventive services task force recommendation statement. *JAMA*, 319, 1901-1913.

Valtonen (1993). Application of cost-benefit thinking in health care. Vaasa: Universitas Wasaensis.

Vasarainen, Malmi, Määttänen, Ruutu, Tammela, Taari, et al. (2013). Effects of prostate cancer screening on health-related quality of life: Results of the Finnish arm of the European randomized screening trial (ERSPC). *Acta Oncologica*, 52, 1615-1621.

Waddington (1986). The perversion of cost-effectiveness. *Radical Community Medicine*, 24, 37-38.

Walsh & Williams (1969). Current issues in cost-benefit analysis. Centre for Administrative Studies Occasional Paper. London: Her Majesty's Stationery Office.

Weatherly, Faria, Berg, Sculpher, O'Neill, Nolan, et al. (2017). Scoping review on social care economic evaluation methods. York, UK: Centre for Health Economics, University of York.

Weinberg (1972). Science and trans-science. *Minerva*, 10, 209-222.

Weinstein (2012). Decision rules for incremental cost-effectiveness analysis. (In Jones (Ed.), *The Elgar Companion to Health Economics* (pp. 505-514). Cheltenham, UK: Edward Elgar).

Weinstein & Zeckhauser (1973). Critical ratios and efficient allocation. *Journal of Public Economics*, 2, 147-157.

Welch & Black (2010). Overdiagnosis in Cancer. *Journal of the National Cancer Institute*, 102, 605-613.

Welte, Feenstra, Jager & Leidl (2004). A Decision Chart for Assessing and Improving the Transferability of Economic Evaluation Results Between Countries. *Pharmacoeconomics*, 22, 857-876.

Wildavsky (1967). The Political Economy of Efficiency: Cost-Benefit Analysis, Systems Analysis, and Program Budgeting. *The Public Interest*, 8, 30-48.

Wildavsky (1969). Rescuing Policy Analysis from PPBS. *Public Administration Review*, 29, 189-202.

Wildavsky (1993). *Speaking truth to power: the art and craft of policy analysis*. New Brunswick (N.J.): Transaction.

Williams (1967). Output Budgeting and the contribution of Micro-Economics to Efficiency in Government? London: HMSO.

Williams (1974). The cost-benefit approach. *British Medical Bulletin*, 30, 252-256.

Williams (1976). Comments by Alan Williams on Bengt Jönsson – Cost-benefit analysis in public health & medical care: Nationalekonomiska Institutionen, Lunds Universitet, Meddelanden.

Williams (1985). Economics of coronary artery bypass grafting. *BMJ*, 291, 326-329.

Williams (1989). The role of health economics. *BMJ*, 299, 679-679.

Williams (1991). Is the QALY a technical solution to a political problem? Of course not! International *Journal of Health Services*, 21, 365-369.

Williams (2004). What could be nicer than NICE? London: The Office of Health Economics (OHE).

Williams (2005). Discovering the QALY, or how Rachel Rosser changed my life. (In Oliver (Ed.), *Personal histories in health research* (pp. 193-206). London: Nuffield Trust for Research and Policy Studies in Health Services).

Wilson & Jungner (1968). Principles and practice of screening for disease. Public Health Papers 34. Geneva: World Health Organisation.

Wiseman (1989a). Economic efficiency and efficient public policy. (In Wiseman (Ed.), *Cost, choice, and political economy* (pp. 171-185). Aldershot: Edward Elgar).

Wiseman (1989b). The way ahead: A new political economy. (In Wiseman (Ed.), *Cost, choice, and political economy* (pp. 265-285). Aldershot, England: Edward Elgar).

Wolff (2004). The 'Efficiency' Illusion. (In Fullbrook (Ed.), *A guide to what's wrong with economics* (pp. 169-175). London: Anthem).

Worthley (1974). PPB: Dead or Alive? *Public Administration Review*, 34, 392-394.

Publications

PUBLICATION I

Health-related quality of life in the Finnish Trial of Screening for Prostate Cancer

Neill Booth, Pekka Rissanen, Teuvo L.J. Tammela, Liisa Määttänen, Kimmo Taari & Anssi Auvinen

European Urology, 65, 39-47

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Platinum Priority – Prostate Cancer

Editorial by Jonathan Bergman and Mark S. Litwin on pp. 48–49 of this issue

Health-Related Quality of Life in the Finnish Trial of Screening for Prostate Cancer

Neill Booth [a,*], *Pekka Rissanen* [a], *Teuvo L.J. Tammela* [b], *Liisa Määttänen* [c], *Kimmo Taari* [d], *Anssi Auvinen* [a]

[a] *School of Health Sciences, University of Tampere, Tampere, Finland;* [b] *Department of Surgery, Tampere University Hospital, and School of Medicine, University of Tampere, Tampere, Finland;* [c] *Mass Screening Registry, Finnish Cancer Registry, Helsinki, Finland;* [d] *Helsinki University Hospital and University of Helsinki, Helsinki, Finland*

Article info

Article history:
Accepted November 18, 2012
Published online ahead of print on November 26, 2012

Keywords:
Quality of life
Prostate cancer
Mass screening
Population-based planning
Randomised controlled trial
Economic value of life

EU ★ ACME
www.eu-acme.org/europeanurology

Please visit www.eu-acme.org/europeanurology to read and answer questions on-line. The EU-ACME credits will then be attributed automatically.

Abstract

Background: Evidence of the potential impact of systematic screening for prostate cancer (PCa) on health-related quality of life (HRQoL) at a population-based level is currently scarce.
Objective: This study aims to quantify the long-term HRQoL impact associated with screening for PCa.
Design, setting, and participants: Postal questionnaire surveys were conducted in 1998, 2000, 2004, and 2011 among men in the Finnish PCa screening trial diagnosed with PCa (total $n = 7011$) and among a random subsample of the trial population ($n = 2200$). In 2011, for example, 1587 responses were received from men with PCa in the screening arm and 1706 from men in the control arm. In addition, from the trial subsample, 549 men in the screening arm and 539 in the control arm provided responses.
Outcome measurements and statistical analysis: Health-state-value scores were compared between the intervention and control arms using three distinct HRQoL measures (15D, EQ-5D, and SF-6D), and statistical significance was assessed using *t* tests. In addition, differences over repeated assessments of HRQoL between groups were evaluated using generalised estimating equations.
Results and limitations: In the 2011 survey, a small but statistically significant difference emerged between the trial arms among men diagnosed with PCa (mean scores, screening vs control arm: 15D: 0.872 vs 0.866, $p = 0.14$; EQ-5D: 0.852 vs 0.831, $p = 0.03$; and SF-6D: 0.763 vs 0.756, $p = 0.06$). Such differences in favour of the screening arm were not found among the sample of men from the trial (15D: 0.889 vs 0.892, $p = 0.62$; EQ-5D: 0.831 vs 0.852, $p = 0.08$; and SF-6D: 0.775 vs 0.777, $p = 0.88$). The slight advantage with screening among men with PCa was reasonably consistent across time in the longitudinal analysis and was strongest among men with early-stage disease.
Conclusions: These results show some long-term HRQoL benefit from screening for men with PCa but suggest little impact overall in the trial population.

* Corresponding author. University of Tampere, School of Health Sciences, Medisiinarinkatu 3, Tampere, FI-33520, Finland. Tel. +358 3 3551 7339; Fax: +358 3 3551 6057.
E-mail address: neill.booth@uta.fi (N. Booth).

1. Introduction

Information on the balance of benefits and adverse effects of screening, which includes quantification of health-related quality of life (HRQoL), is an important aspect of research and relevant for both clinical and health-care decision making [1,2]. There are many ways of measuring HRQoL aimed at capturing the overall effects of an intervention on

0302-2838/$ – see back matter
http://dx.doi.org/10.1016/j.eururo.2012.11.041

functional capacity, well-being, and health [3]. One approach, commonly advocated by health economists, uses utility values for health states (ie, health-state-value scores), which can subsequently be used in the calculation of quality-adjusted life-years (QALYs) [4].

This study presents the results of HRQoL surveys carried out as part of an economic evaluation of the Finnish population-based randomised controlled trial of screening for prostate cancer (FinRSPC) [5,6]. The trial has secondary objectives of evaluating HRQoL and cost effectiveness. We used three generic measures of HRQoL chosen in line with the public health focus of the trial to facilitate future cost-effectiveness estimations [7]. The mean health-state-value scores are compared between the trial arms, separately among men diagnosed with prostate cancer (PCa) and a random subsample of the target population.

2. Methods

2.1. Trial background

The target population of the FinRSPC consisted of men born in from 1929 to 1944 who resided in the Helsinki or Tampere region during the recruitment period (1996–1999) identified from the population registry (n = 80 458). An exclusion criterion was a diagnosis of PCa before the date of randomisation (obtained from the Finnish Cancer Registry). Men who were known to have died, had prohibited use of their information for research, or had moved outside the study area between randomisation and the date of mailing were also considered ineligible (n = 314).

The men in the screening arm were sent an invitation letter to screening that also explained the study's purpose and procedures, as well as general information on PCa. Of the 31 866 men contactable at the time of invitation, 1597 were not sent an invitation because of logistical problems. The men were invited for second and third screening rounds

Table 1 – Background information for the study participants in the two groups of men*

Random sample of all men in the trial						
	Screening arm			Control arm		
	n	Median	IQR	*n*	Median	IQR
Domicile						
Helsinki	573			568		
Tampere	180			185		
Age, yr						
1998	740	58.87	62.80–54.99	733	58.92	62.99–54.97
1999	748	60.09	64.07–56.18	752	60.11	64.19–56.17
2003	683	63.71	67.67–59.85	690	63.77	67.80–59.83
2011	549	71.63	75.70–67.79	539	71.61	75.66–67.83
Socioeconomic status†						
Upper	179	–	–	174	–	–
Middle	87	–	–	92	–	–
Lower	238	–	–	227	–	–
Men diagnosed with PCa						
Domicile						
Helsinki	2021	–	–	1670	–	–
Tampere	862	–	–	700	–	–
Age, yr						
1998	146	64.34	67.65–60.31	34	63.81	67.50–60.71
1999	260	65.30	68.51–61.41	94	64.90	68.68–61.34
2003	890	67.46	71.03–63.95	536	67.67	70.83–63.71
2011	1587	73.56	77.08–70.04	1706	73.68	77.54–70.13
Time since diagnosis, yr						
1998	145	1.61	2.08–1.17	34	1.78	2.18–1.39
1999	257	2.89	3.50–2.22	94	2.53	3.34–1.83
2003	890	3.63	5.24–1.94	536	2.92	4.22–1.49
2011	1313	8.00	10.81–5.57	1431	6.70	9.01–5.02
Tumour stage‡						
T1–2	2628	–	–	1991	–	–
T3–4	255	–	–	379	–	–
Socioeconomic status						
Upper	807	–	–	738	–	–
Middle	301	–	–	300	–	–
Lower	1093	–	–	890	–	–

IQR = interquartile range; PCa = prostate cancer.

* The random sample of all men in the trial (formed in 1998) and the men diagnosed with prostate cancer (with new cases recruited at each round).

† The levels of socioeconomic status are defined in accordance with a classification developed by Statistics Finland [47].

‡ TNM T cancer stage classification category at the time of diagnosis with PCa [48].

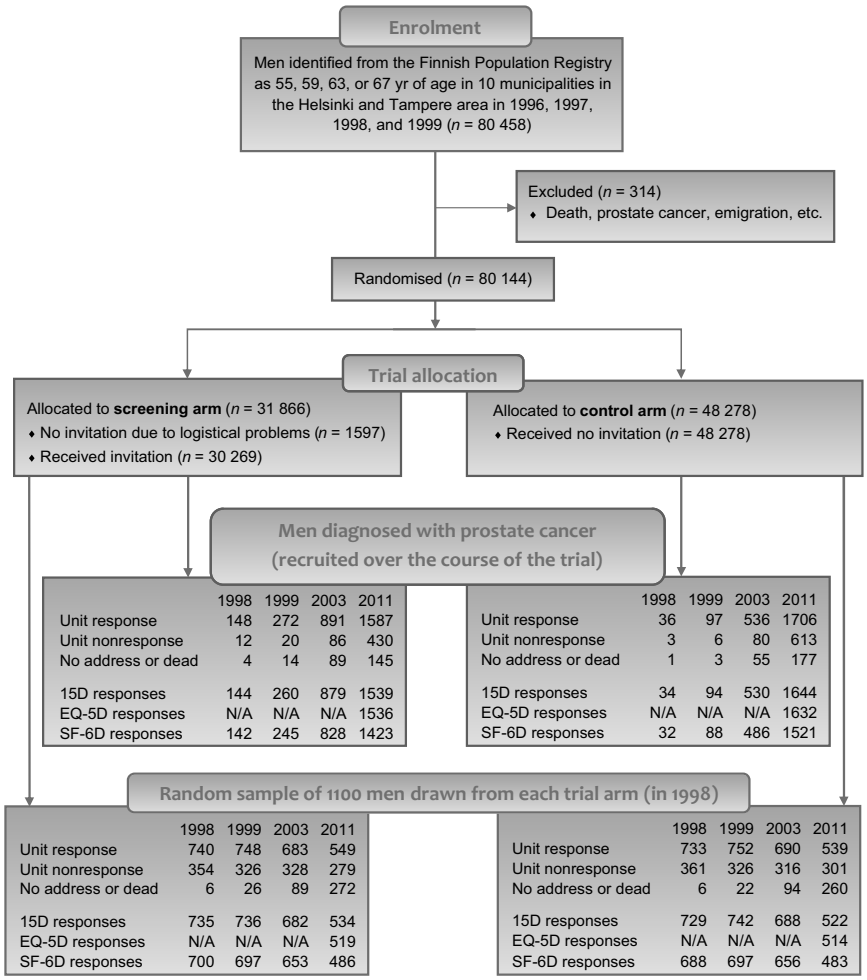

Fig. 1 – Flowchart illustrating the data collection process. N/A = not applicable.

similarly, 4 and 8 yr after the first screen (although men >71 yr of age were no longer invited, so those age 67 at the initial screen were invited only twice). Information on screening invitations and attendance, screening (prostate-specific antigen [PSA]) and diagnostic test results, clinical information on PCa, and death were systematically recorded in the trial database [8]. The control group received no systematic invitation and were not contacted as part of the FinRSPC [9], except for those in the trial population subsample who received postal questionnaires.

2.2. Study materials

Two groups of men from the trial received postal questionnaires concerning their HRQoL (Table 1 and Fig. 1). One group consisted of *men diagnosed with PCa* by 1998, 1999, 2003, and 2011 who were identified from the trial database and/or the Finnish Cancer Registry prior to each of those time points. The second group of men was randomly sampled from the men inducted to the FinRSPC in 1998, hereafter referred to as the *trial subsample*. This trial subsample consisted of 2200 men: 1100 from the screening arm and 1100 men from the control arm. These same 2200 men were approached four times consecutively (in 1998, 1999, 2003, and 2011), although emigrated, deceased, and otherwise untraceable men were not eligible. They were all free of PCa at baseline, but 108 in the screening arm and 84 in the control arm were subsequently diagnosed with the disease.

Data from the trial database and postal questionnaires for each individual were interlinked using the Finnish system of unique personal identity codes. Men with PCa gave consent to link questionnaire responses with the trial database and cancer registry files in all rounds of the questionnaire. Men in the 1998 random sample of all participants (ie, in the trial subsample) consented in the latest questionnaire. Investigators received authorisation from the Finnish National Institute of Health and Welfare to use the questionnaire responses from the men in the trial subsample during the other rounds. The study plan for this study, as part of the economic evaluation, was reviewed by the Tampere University Hospital ethics committee (reference number R05053).

The HRQoL questionnaires used in this study are the RAND 36-Item Short Form Health Survey [10], which can be used to produce the SF-6D measure [11]; the 15D health state description system [12]; and the EQ-5D instrument (with the UK-TTO scoring system) [13]. Further details on each of these measures can be obtained from online resources [14–16]. We used the SF-36 because it is one of the most widely used generic HRQoL questionnaires and the 15D measure because it is one of the most comprehensive of the generic measures. In 2011, we also used the EQ-5D because it is reported to be responsive [17] and one of the simplest and most commonly used measures [18]. In line with standard use of these measures, no attempt was made to correct for either EQ-5D or SF-6D item nonresponse, but in cases of missing values for one to three dimensions of the 15D, replacement values were imputed through linear regression analysis using the remaining dimensions and age as explanatory variables [19]. All these HRQoL measures, wherein higher health-state-value scores are assumed to represent better outcomes, can produce mean scores at group levels required to construct QALYs, which can be used to provide information for health care policy decisions [20].

This study comprises both cross-sectional and longitudinal information. Differences between the trial arms in the mean scores from the 2011 questionnaires were analysed using two sample mean comparison tests. The HRQoL scores among men diagnosed with PCa were compared with men from the trial subsample as the reference [21]. Finally, differences in scores for both groups of men over time (longitudinal data) were assessed using generalised estimating equations (GEEs). This method can take into account correlations between intraindividual observations in repeated assessments and makes use of each response even if a man only responded to one questionnaire. The GEE method was used to investigate the covariates of socioeconomic status, domicile, and age group in the four rounds of postal questionnaires. For the analysis of dropout (ie, nonresponse to the questionnaire to men with PCa), we used logistic regression; in addition to investigating the association of nonresponse with the trial arm, we included information on hospital episodes during the study period and tumour stage at PCa (t tests), available from the trial database.

3. Results

Between 1998 and 2011, HRQoL questionnaires with responses were received from 5516 men in total (ie, from 79.4% of men who were alive and whose addresses were available during the four questionnaire rounds). A maximum of four questionnaires were sent to the men with PCa and four sent to the trial subsample. The total number of responses from men with PCa during the whole study period was 2898 in the screening arm and 2375 in the control arm. The percentage ratio of the responders to the total numbers of cases was 78.4% in the screening arm and 71.7% in the control arm (see also Fig. 1 [22]). Health-state-value scores were able to be calculated for most men who had returned their questionnaire; for example, HRQoL scores from the 15D measure could be calculated for 94% of respondents. The ability to calculate scores was similar, although slightly lower, for the other HRQoL measures, the EQ-5D and the SF-6D (Fig. 1).

3.1. Cross-sectional analyses

In 2011, the differences in mean HRQoL scores between the trial arms were statistically significant only with the EQ-5D measure and only among men with PCa (Fig. 2). The difference in the EQ-5D score was an increment of 0.016 in favour of the screening arm ($p = 0.017$); however, the 15D also showed a smaller non-statistically-significant difference in HRQoL score ($p = 0.007$), for example. Regression analysis confirmed that these results were also robust to adjustment for time since diagnosis and PCa stage (not shown).

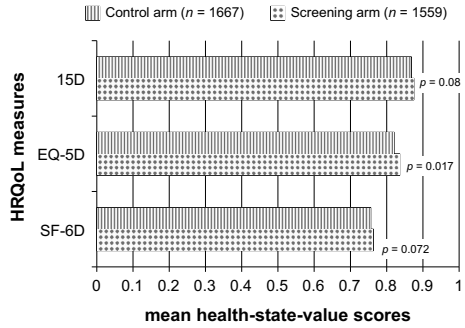

Fig. 2 – Differences in mean health-related quality-of-life scores in 2011 between the Finnish prostate cancer (PCa) screening trial arms for men diagnosed with PCa (p values refer to two-sided t tests).

For men from each arm of the trial subsample, no statistically-significant differences were observed for any of the measures (Fig. 3). Although when men diagnosed with PCa were excluded from the analysis of the trial subsample, scores were higher in the control arm (15D: 0.890 vs 0.895, $p = 0.55$; EQ-5D: 0.830 vs 0.857, $p = 0.04$; and SF-6D: 0.777 vs 0.779, $p = 0.80$, not shown).

In the 2011 survey, the decrement in the mean HRQoL scores of men with PCa relative to the trial subsample was slightly more pronounced in the control arm than in the screening arm (Fig. 4). In the control arm, the mean scores of all HRQoL measures for men with PCa were lower than those for men from the trial subsample, whereas in the screening arm men with stage T1 or T2 PCa had higher or similar mean HRQoL scores than men from the trial subsample (Fig. 5). Among men with stage T3 or T4 PCa in both arms, a decrement in mean HRQoL scores was

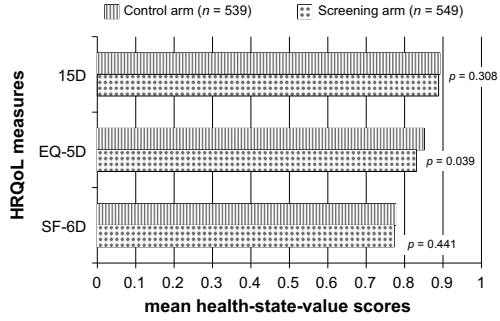

Fig. 3 – Differences in mean health-related quality-of-life scores in 2011 between the Finnish prostate cancer screening trial arms in a random sample of trial participants (p values refer to two-sided t tests).

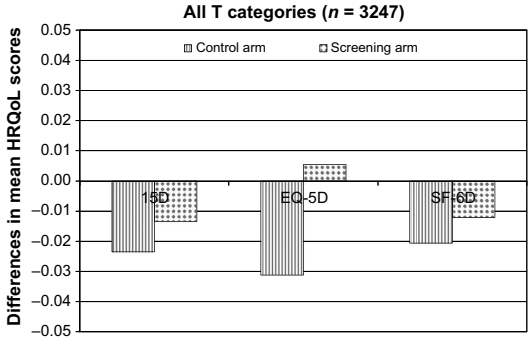

Fig. 4 – Differences in mean health-related quality-of-life (HRQoL) scores in 2011 between men diagnosed with prostate cancer and a random sample of men in each trial arm.

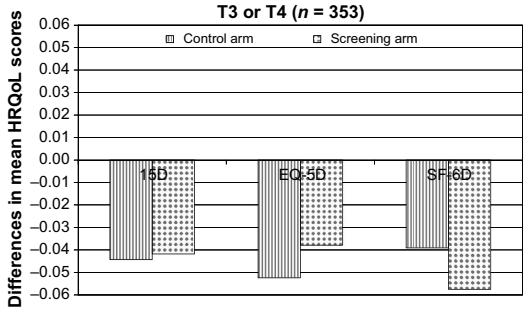

Fig. 6 – Differences in mean health-related quality-of-life (HRQoL) scores in 2011 between men diagnosed with advanced prostate cancer and a random sample of all men in the respective trial arm.

apparent for men in the control arm and the screening arm, in comparison with the reference group of men from the trial subsample (Fig. 6).

3.2. Longitudinal analysis

In the longitudinal data analysis for men from the trial subsample, using data collected between 1998 and 2011, only age and socioeconomic status were statistically significant determinants of the 15D score (Table 2), but no such differences were found between trial arms or localities (domicile in the Helsinki or Tampere area). For men with PCa, the mean 15D scores in all surveys for men in the screening arm were higher (by an increment of 0.01) than in the control arm after adjustment for age, domicile, and socioeconomic status (see Table 3 and Fig. 7). Statistically significant reductions in 15D scores were also associated with lower socioeconomic status, residence in the Tampere area, and increased age.

Table 4 extends the analysis to compare the HRQoL scores of men with screen-positive PCa to those of men with PCa in the control arm. The regression results were very similar to those in Table 3, but the screen-positive men in the screening arm had a larger increment, 0.016, in their mean 15D scores.

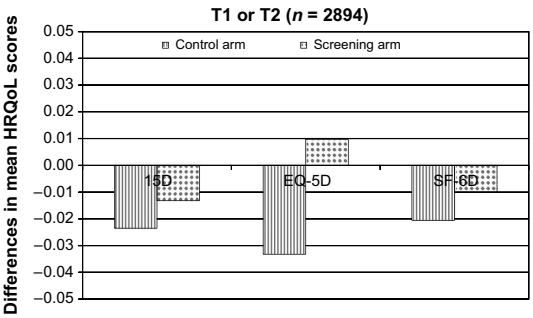

Fig. 5 – Differences in mean health-related quality-of-life (HRQoL) scores in 2011 between men diagnosed with organ-confined prostate cancer and a random sample of all men in each trial arm.

3.3. Nonresponse analysis

In the dropout analysis for the 2011 survey among men with PCa, no statistically significant difference in nonparticipation was observed between the trial arms (Table 5). Being older age and advanced cancer (stage T3–4 compared with T1–2) were associated with nonresponse, and a lower risk of nonresponse was associated with an episode of hospital care during the study period. Similar analysis was undertaken for the 2003 survey, with comparable results (not shown).

4. Discussion

The public-health impact of PCa screening potentially involves mortality reduction, changes in the use of health-care resources, and an impact on morbidity by reducing advanced PCa, but also adverse effects of screening. Here we used several generic measures of health

Table 2 – The effect of trial arm, domicile, socioeconomic status, and age on health-related quality of life in a random sample of all trial participants*

	Coefficient	Robust standard error	*p* value
Screening arm (control arm as reference)	0.000	0.004	0.934
Tampere (Helsinki as reference)	−0.002	0.005	0.724
Socioeconomic status (highest socioeconomic group as reference)			
Middle	−0.008	0.006	0.188
Lowest	−0.039	0.005	0.000
Age, yr	−0.002	0.000	0.000
Constant	1.083	0.013	0.000
Observations = 5315			

* Longitudinal analysis of 15D scores in the Finnish prostate cancer screening trial between 1998 and 2011 with generalised estimating equations.

Table 3 – The effects of trial arm, domicile, socioeconomic status, and age on health-related quality of life among men with prostate cancer*

	Coefficient	Robust standard error	*p* value
Screening arm (control arm as reference)	0.010	0.003	0.002
Tampere (Helsinki as reference)	−0.018	0.004	0.000
Socioeconomic status (highest socioeconomic group as reference)			
Middle	−0.014	0.005	0.003
Lowest	−0.021	0.004	0.000
Age, yr	0.004	0.000	0.000
Constant	0.986	0.009	0.000
Observations = 4054			

* Longitudinal analysis of 15D scores in prostate cancer (PCa) in 1998 and 2011 in the Finnish PCa screening trial with generalised estimating equations.

status to evaluate the long-term HRQoL effects of screening on all men in a screening trial as well as those diagnosed with PCa. Our results suggest a small overall benefit among men with PCa, although a statistically significant difference was not observed at all time points or for all measures. The effect size (ie, the difference observed in HRQoL scores) is small, both in absolute terms and relative to the variance of the scores. No clear differences were found overall in a random sample of men in the trial, although in a subgroup analysis, a slightly lower overall score was seen if men with PCa were excluded from the analysis.

We did not attempt to compare the results given by the three different HRQoL measures but rather provided a description of the HRQoL scores. However, of the three measures, the 15D had the smallest variance and the highest HRQoL scores, the EQ-5D had the largest variance

Table 4 – The effects of domicile, socioeconomic status, and age on health-related quality of life among men diagnosed with prostate cancer restricted to screen-positive men in the screening arm*

	Coefficient	Robust standard error	*p* value
Positive screen result (relative to the men with PCa in the control arm)	0.016	0.003	0.000
Tampere (Helsinki as reference)	−0.018	0.004	0.000
Socioeconomic status (highest socioeconomic group as reference)			
Middle	−0.014	0.005	0.003
Lowest	−0.022	0.004	0.000
Age, yr	0.004	0.000	0.000
Constant	0.985	0.009	0.000
Observations = 4054			

PCa = prostate cancer.
* Longitudinal analysis of 15D scores in four questionnaire surveys of PCa carried out between 1998 and 2011 with generalised estimating equations.

Table 5 – Dropout analysis: relationship of trial arm, prior hospitalisation, age, and tumour stage to participation*

	Odds ratio	Robust standard error	*p* value
Screening arm (control arm as reference)	0.995	0.067	0.942
Hospital care (no recorded hospital episodes as reference)	1.359	0.098	0.000
Age, yr	0.972	0.008	0.000
Advanced stage PCa (T3–4 relative to organ-confined T1–2)	0.755	0.076	0.005
Constant	19.432	11.586	0.000
Observations = 4526			

PCa = prostate cancer.
* In the postal questionnaire survey in 2011 among men diagnosed with PCa.

and scores slightly lower than the 15D, and the SF-6D measure had variance slightly larger than 15D and produced the lowest HRQoL scores. All three measures agreed on the direction of the effects, but the absolute size of the differences, and the most statistically significant results, were found using the EQ-5D scoring system. This is consistent with some recent research comparing these measures [23,24].

No improvement was found overall for men in the screening arm of the trial subsample, which can be expected because a cancer screening programme is unlikely to yield large long-term benefits among men not diagnosed with PCa. However, in a subgroup analysis, for those men in the trial subsample free of PCa in 2011, an overall benefit was suggested for men in the control arm. It is unclear if this represents a negative impact of screening because the short-term impact of attending screening on HRQoL was previously often found to be minor and transient [25]. In the trial subsample, it is likely that the relatively higher mean scores of men with PCa in the screening arm helped to raise the overall mean scores for men in the screening arm, thus negating any small increase in the overall mean scores for men in the control arm. For these reasons, the existence of any long-term detriment overall in the screening arm would need substantiation through further study. One repeated finding from our analyses was lower HRQoL associated with more advanced age and lower socioeconomic status.

Our material was based on a large population-based trial and provides evidence about the impact of mass screening on the general population. Here, even small differences, below those conventionally regarded as clinically relevant, may be of public health interest [26–28].

4.1. *Strengths of the study*

Although research on PCa patients' HRQoL has been conducted [29–32], few studies have examined the effect of screening on HRQoL, and even fewer have examined data from a randomised screening trial [33–35]. The setting of a randomised trial is a major advantage and regarded as the gold standard for evaluating medical and public health

interventions. Our primary analysis followed the intention-to-screen principle, rather than applying screening-received analysis, to improve validity and generalisability [6,36]. This means that, in our analysis, nonparticipants were included in the screening arm and men having opportunistic PSA testing were included in the control arm, which is likely to give smaller effect than comparing screened and unscreened men but a more realistic estimate of the impact of population-based screening. We did not control for the effect of treatment or stage of PCa because it is likely affected by screening and hence would lead to overadjustment.

We evaluated the HRQoL over a 13-yr period: during the first, second, and third screening round, as well as after completion of the intervention (Fig. 7). Therefore, the results complement earlier work focusing on the short-term impact of the screening process, diagnostic procedures, and treatment [29–31,34,35].

4.2. Limitations of the study

Postal administration was used to reach both men diagnosed with PCa and men not otherwise contacted as part of the trial in a similar manner, with response proportions in both groups relatively high [37]. This Finnish patient cohort is particularly homogeneous in terms of ethnic background and socioeconomic status. Therefore, it may be inappropriate to generalise these findings to heterogeneous populations with wider variations in ethnicity and socioeconomic status. One of the main limitations of this study is that we found some indication of selection bias, with nonrespondents being older and more likely to have advanced stage PCa at diagnosis, but nonresponse was not associated with the trial arm. Therefore, we did not use imputation methods for unit nonresponse: The assumptions involved are necessarily arbitrary [38]. However, because of the differences in age distribution, especially those between the group of men with PCa and the group of men from the trial subsample, the differences in mean health-state-value scores in 2011, as shown in Figures 4–6, may not be amenable to simplistic interpretation. If

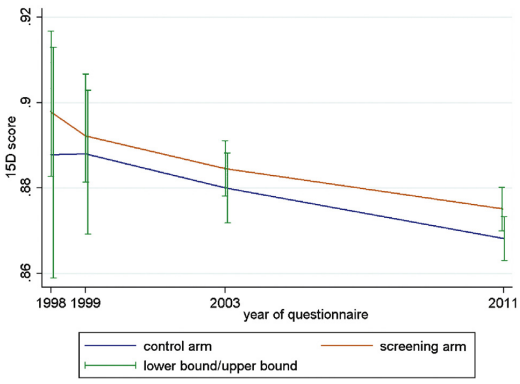

Fig. 7 – Longitudinal differences in mean 15D scores (with 95% confidence intervals) for men diagnosed with prostate cancer (PCa) by arm in the Finnish PCa screening trial.

these differences in age distribution were not present and selection bias was not a problem, it might be appropriate to estimate the partial impact of screening in terms of morbidity-related QALYs. Such an estimation involves multiplying the mean differences between HRQoL scores for men with PCa and those from the trial subsample in each arm (Fig. 4) by the cumulative incidence of PCa in each arm (as per Schröder et al. [6]). In our study, the appropriate comparator for the men with PCa was taken to be all men in their respective arm of the trial subsample, rather than just those without a PCa diagnosis. As an example of such a simplistic calculation, using the 15D scores collected in 2011, the estimate would be a gain of less than one morbidity-related QALY per 1000 trial participants in that year.

In general, our results may also reflect opportunistic PSA testing in the control arm—increasing levels of contamination within the trial—which may reduce the relative benefit found from systematic mass screening [39]. In addition, lead-time bias, length bias, and overdiagnosis are likely to lead to larger numbers of small early cancers being detected at a younger age in the screening arm, which may overestimate the screening benefit. However, this would not affect the comparison of all men in the trial (when men with PCa are included).

We used three validated Finnish-language HRQoL questionnaires that can produce health-state-value scores as indicators of overall health impact. None of these are disease-specific measures, and it remains somewhat unclear how comprehensively they can assess the adverse health outcomes typical for PCa. It also remains unclear what results would have been obtained if a disease-specific measure, such as the UCLA Prostate Cancer Index, had been used [40]. Despite the fact that they have been developed for the evaluation of key aspects of disease and ill health limiting everyday life, generic HRQoL measures are less likely to react to small or insidious changes in health status than major or acute changes. Furthermore, the impact of PCa may well be masked by other medical conditions. Generic HRQoL measures such as those used here may not be wholly suitable for describing the effects of palliative care [41] or, for example, the effects of illness on family members. Assessment of the challenges posed by response shift [42,43], scale recalibration, or adaptation [44] was not undertaken here, and, of course, the HRQoL measures may not faithfully represent the societal value of changes in health status [45]. Instead, each HRQoL measure generates alternative valuations of respondents' subjective responses to systematic questions addressing their health status. Further, these measures are aimed at detecting differences in group-level HRQoL, rather than clinically important changes at the individual patient level and hence are rather more suitable for guiding policy than for guiding clinical decision making concerning individual patients [46]. These HRQoL results should be set in the appropriate context; namely, judgements concerning the overall merits of screening are generally a combination of mortality effects, morbidity estimates (such as the HRQoL information presented here), and disease-specific traits, as well as other health-related, economic, and cultural values. Regardless of

the rather small effect observed, subsequent application of these HRQoL scores to describe health state utilities in cost-effectiveness analyses remains a possibility.

5. Conclusions

This study shows a small advantage in mean HRQoL scores for the screening arm over the control arm for men diagnosed with PCa. These differences were small and not detected by all of the generic indicators at all times in the course of the 13-yr follow-up. Using these HRQoL measures, this study provides little evidence that mean health-state-value scores differed markedly between the trial arms for the trial population overall.

Author contributions: Neill Booth had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Rissanen, Auvinen, Booth.
Acquisition of data: Määttänen, Booth, Auvinen, Rissanen, Tammela.
Analysis and interpretation of data: Booth, Auvinen, Rissanen.
Drafting of the manuscript: Booth.
Critical revision of the manuscript for important intellectual content: Auvinen, Booth, Taari, Tammela, Rissanen.
Statistical analysis: Booth, Auvinen.
Obtaining funding: Rissanen, Tammela, Auvinen, Booth.
Administrative, technical, or material support: None.
Supervision: Rissanen, Auvinen.
Other (specify): None.

Financial disclosures: Neill Booth certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Kimmo Taari has received speaker honoraria from GSK; consultant honoraria from Astellas, GSK, Ferring, and Amgen; and taken part in congresses with support from Sanofi-Aventis, Pfizer, and Astellas. Teuvo Tammela has consulted for AstraZeneca, GlaxoSmithKline, Orion Pharma, and Pfizer, and received honoraria from Astellas, GlaxoSmithKline, Leiras, and Pfizer. Neill Booth, Pekka Rissanen, Anssi Auvinen, and Liisa Määttänen have nothing to disclose.

Funding/Support and role of the sponsor: This study was funded by unrestricted grants from the Competitive Research Funding of the Tampere University Hospital (Grants 9F100, 9L085, 9E089 and 9N064). The work of Neill Booth was also supported by the Yrjö Jahnsson Foundation (Grant 6213), FinOHTA, and Doctoral Programs in Public Health (DPPH) funding. Also aiding greatly in the completion of the paper was a study visit to the PCMD Health Economics Group, University of Exeter, UK.

Acknowledgment statement: The authors would like to thank all those who have helped in collecting the data, especially the men in the trial who took the time and made the effort to respond to the questionnaires.

References

[1] Heidenreich A, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. Eur Urol 2011;59:61–71.
[2] Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert Rev Pharmacoecon Outcomes Res 2010;10:553–66.
[3] Mosteller F. Implications of measures of quality of life for policy development. J Chronic Dis 1987;40:645–50.
[4] Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986;5:1–30.
[5] Hakama M, Aro J, Auvinen A, et al. Randomized screening trial for prostate cancer in Finland. Eur Urol 2001;39(Suppl 4):32.
[6] Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366:981–90.
[7] Drummond M, O'Brien B. Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. Health Econ 1993;2:205–12.
[8] Määttänen L, Auvinen A, Stenman UH, et al. European randomized study of prostate cancer screening: first-year results of the Finnish trial. Br J Cancer 1999;79:1210–4.
[9] Määttänen L, Auvinen A, Stenman UH, et al. Three-year results of the Finnish prostate cancer screening trial. J Natl Cancer Inst 2001;93:552–3.
[10] Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. Health Econ 1993;2:217–27.
[11] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.
[12] Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;33:328–36.
[13] The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
[14] SF-6D. Health Economics and Decision Science section of the University of Sheffield Web site. http://www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d. Accessed November 3, 2012.
[15] Sintonen H. Information concerning the 15D instrument, its properties and usage. 15D Web site. http://www.15d-instrument.net/. Accessed November 3, 2012.
[16] EuroQol home page. EuroQol Group Web site. http://www.euroqol.org/. Accessed November 3, 2012.
[17] Krabbe PF, Peerenboom L, Langenhoff BS, Ruers TJ. Responsiveness of the generic EQ-5D summary measure compared to the disease-specific EORTC QLQ C-30. Qual Life Res 2004;13:1247–53.
[18] Cookson R, Culyer A. Measuring overall population health—the use and abuse of QALYs. In: Killoran A, Kelly MP, editors. Evidence-based public health: effectiveness and efficiency. Oxford, UK: Oxford University Press; 2010. p. 148–68.
[19] Sintonen H. Replacing missing data. 15D Web site. http://www.15d-instrument.net/rmd. Accessed June 5, 2012.
[20] Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000;320:1197–200.
[21] Bowling A. Research methods in health: investigating health and health services. Maidenhead, UK: Open University Press; 2009.
[22] Schulz KF, Altman DG, Moher D., CONSORT group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Ann Intern Med 2010;152:726–32.
[23] Kontodimopoulos N, Pappa E, Chadjiapostolou Z, Arvanitaki E, Papadopoulos AA, Niakas D. Comparing the sensitivity of EQ-5D, SF-6D and 15D utilities to the specific effect of diabetic complications. Eur J Health Econ 2012;13:111–20.
[24] Sørensen J, Linde L, Østergaard M, Hetland ML. Quality-adjusted life expectancies in patients with rheumatoid arthritis—comparison of index scores from EQ-5D, 15D, and SF-6D. Value Health 2012;15:334–9.
[25] Cullen J, Schwartz MD, Lawrence WF, Selby JV, Mandelblatt JS. Short-term impact of cancer prevention and screening activities on quality of life. J Clin Oncol 2004;22:943–52.
[26] King MT. A point of minimal important difference (MID): a critique of terminology and methods. Expert Rev Pharmacoecon Outcomes Res 2011;11:171–84.

[27] Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. J Clin Oncol 1998;16:139–44.
[28] Glass GV, McGaw B, Smith ML. Meta-analysis in social research. Beverly Hills, CA: Sage; 1984.
[29] Namiki S, Arai Y. Health-related quality of life in men with localized prostate cancer. Int J Urol 2010;17:125–38.
[30] McNaughton-Collins M, Walker-Corkery E, Barry MJ. Health-related quality of life, satisfaction, and economic outcome measures in studies of prostate cancer screening and treatment, 1990–2000. J Natl Cancer Inst Monogr 2004;78–101.
[31] Penson DF, Rossignol M, Sartor AO, Scardino PT, Abenhaim LL. Prostate cancer: epidemiology and health-related quality of life. Urology 2008;72:S3–11.
[32] Heijnsdijk EAM, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med 2012;367:595–605.
[33] Taylor KL, Luta G, Miller AB, et al. Long-term disease-specific functioning among prostate cancer survivors and noncancer controls in the prostate, lung, colorectal, and ovarian cancer screening trial. J Clin Oncol 2012;30:2768–75.
[34] Madalinska JB, Essink-Bot ML, de Koning HJ, Kirkels WJ, van der Maas PJ, Schroder FH. Health-related quality of life in patients with screen-detected versus clinically diagnosed prostate cancer preceding primary treatment. Prostate 2001;46:87–97.
[35] Korfage IJ, Essink-Bot ML, Borsboom GJ, et al. Five-year follow-up of health-related quality of life after primary treatment of localized prostate cancer. Int J Cancer 2005;116:291–6.
[36] Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research. Int J Epidemiol 1992;21: 837–41.
[37] Johnson TP, Wislar JS. Response rates and nonresponse errors in surveys. JAMA 2012;307:1805–6.
[38] Rubin DB. Multiple imputation for nonresponse in surveys. Hoboken, NJ: Wiley Interscience; 2004.
[39] Ciatto S, Zappa M, Villers A, Paez A, Otto SJ, Auvinen A. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. BJU Int 2003;92(Suppl 2):97–100.
[40] Litwin MS, Hays RD, Fink A, et al. Quality-of-life outcomes in men treated for localized prostate cancer. JAMA 1995;273:129–35.
[41] Hughes J. Palliative care and the QALY problem. Health Care Anal 2005;13:289–301.
[42] Sprangers M, Schwartz C. Do not throw out the baby with the bath water: build on current approaches to realize conceptual clarity. Response to Ubel, Peeters, and Smith. Qual Life Res 2010;19:477–9.
[43] Sprangers MA. Response-shift bias: a challenge to the assessment of patients' quality of life in cancer clinical trials. Cancer Treat Rev 1996;22(Suppl A):55–62.
[44] Ubel P, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. Qual Life Res 2010;19:465–71.
[45] Green A, Barker C. Priority setting and economic appraisal: whose priorities—the community or the economist? Soc Sci Med 1988;26:919–29.
[46] Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Econ 2002;11:1–8.
[47] Statistics Finland. Classification of socio-economic groups 1989. Helsinki, Finland: Statistics Finland; 1989.
[48] Sobin LH, Wittekind C, Gospodarowicz MK. TNM classification of malignant tumours. Oxford, UK: Wiley-Blackwell; 2010.

PUBLICATION II

Costs of screening for prostate cancer: Evidence from the Finnish Randomised Study of Screening for Prostate Cancer after 20-year follow-up using register data

Neill Booth, Pekka Rissanen, Teuvo L.J. Tammela, Kimmo Taari, Kirsi Talala & Anssi Auvinen

European Journal of Cancer, 93, 108-118

European Journal of Cancer 93 (2018) 108–118

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Costs of screening for prostate cancer: Evidence from the Finnish Randomised Study of Screening for Prostate Cancer after 20-year follow-up using register data

Neill Booth [a,*], Pekka Rissanen [a], Teuvo L.J. Tammela [b,c], Kimmo Taari [d], Kirsi Talala [e], Anssi Auvinen [a]

[a] *Faculty of Social Sciences (Health Sciences), University of Tampere, FI-33014 Tampere, Finland*
[b] *Department of Urology, Tampere University Hospital, FI-33521 Tampere, Finland*
[c] *Faculty of Medicine and Life Sciences, University of Tampere, FI-33014 Tampere, Finland*
[d] *Department of Urology, University of Helsinki and Helsinki University Hospital, FI-00029 Helsinki, Finland*
[e] *Finnish Cancer Registry, FI-00130 Helsinki, Finland*

Received 5 December 2017; received in revised form 18 January 2018; accepted 30 January 2018

KEYWORDS
Prostatic neoplasms;
Prostate-specific antigen;
Mass screening;
Randomised controlled trial;
Costs and cost analysis

Abstract ***Objectives:*** Few empirical analyses of the impact of organised prostate cancer (PCa) screening on healthcare costs exist, despite cost-related information often being considered as a prerequisite to informed screening decisions. Therefore, we estimate the differences in register-based costs of publicly funded healthcare in the two arms of the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC) after 20 years.
Methods: We obtained individual-level register data on prescription medications, as well as inpatient and outpatient care, to estimate healthcare costs for 80,149 men during the first 20 years of the FinRSPC. We compared healthcare costs for the men in each trial arm and performed statistical analysis.
Results: For all men diagnosed with PCa during the 20-year observation period, mean PCa-related costs appeared to be around 10% lower in the screening arm (SA). Mean all-cause healthcare costs for these men were also lower in the SA, but differences were smaller than for PCa-related costs alone, and no longer statistically significant. For men dying from PCa, although the difference was not statistically significant, mean all-cause healthcare costs were around 10% higher. When analysis included all observations, cumulative costs were slightly higher in the CA; however, after excluding extreme values, cumulative costs were slightly higher in the SA.

* *Corresponding author*:
E-mail address: Neill.Booth@uta.fi (N. Booth).

https://doi.org/10.1016/j.ejca.2018.01.111

Conclusions: No major cost impacts due to screening were apparent, but the FinRSPC's 20-year follow-up period is too short to provide definitive evidence at this stage. Longer term follow-up will be required to be better informed about the costs of, or savings from, introducing mass PCa screening.

1. Introduction

Although there is some evidence of the effectiveness of organised screening in reducing prostate cancer (PCa) mortality [1], there has been a dearth of published empirical analyses of the actual impact of such mass screening on healthcare costs in real-world settings. Prostate-specific antigen (PSA)–based screening potentially provides a means of altering the clinical course of the PCa and thereby improving prognosis and outcomes [2]. However, a presumption is often made that early intervention will reduce overall healthcare costs ([2–4]), and this presumption should be assessed, ideally through a pragmatic randomised controlled trial (RCT) ([5,6]). The primary objective of this analysis is to compare register-based healthcare cost estimates between the two arms of the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC), primarily using intention-to-screen (ITS)—analysis after a maximum of 20 years of follow-up.

2. Methods

2.1. Participants and intervention

Although the European Randomized study of Screening for Prostate Cancer (ERSPC) offers comparable data from each participating centre on outcome measures such as PCa mortality [1], it is unlikely that the ERSPC can offer comparable data on healthcare costs, as costs are known to be dependent on the healthcare system in question [7]. Given such differences in cost accounting and costs even within Europe, and the well-established registers of healthcare cost-related information in Finland, our study is restricted to the FinRSPC, which contributes the largest number of trial participants to the ERSPC. The analysis of healthcare costs presented here is carried out as part of the FinRSPC, the primary objective of which is to investigate the impact of mass PSA-screening on PCa mortality [8]. Secondary objectives of the FinRSPC include the investigation of the trial's impact on costs and health-related quality of life, and then the combination of these sources of information to provide information on cost-effectiveness [9]. The target population of the FinRSPC was selected from the Finnish population registry and consists of men born in 1929–1944 and residing in the Helsinki or Tampere region during the recruitment period (1996–99, total randomised n = 80,458). The main exclusion criterion was PCa-diagnosis before the date of randomisation (this information was obtained from the Finnish Cancer Registry, [FCR]). Further details about the study design can be obtained from Booth *et al.* [10]. The men in the screening group (screening arm, SA) were invited to the screening test (serum PSA) at a local clinic. The men in the reference group (control arm, CA) received no invitation as part of the trial.

3. Materials and analytical methods

The research protocol for the present study was approved by Finnish data-protection authorities and by the National Institute for Health and Welfare (THL). The protocol was also reviewed by the Tampere University Hospital Ethics committee (reference number R05053). After receiving study approval, we were permitted to collate and link the data supplied by a number of registries to the FinRSPC database, using each man's unique Finnish personal identity code for retrieval. This study was undertaken in close co-operation with the FCR, with resources and expertise from the FCR helping to create, maintain, and improve the FinRSPC trial database and its links with the FCR's cancer register [11]. The main data sources used in this study are described in the Appendix: these are the FinRSPC trial database, the Care Register for Health Care (CRHC) and the prescription-medicine reimbursement register (PMRR). The costs of the screening intervention have been estimated to be approximately 50 Euros per screen (including the organisation of the invitation, the drawing of the blood sample and the PSA determinations), and this figure is used in all analyses. All total or average Euro amounts we report in our results are rounded to the nearest 100 Euros, as this gives a suitable level of precision for these cost estimates. The information on screening and healthcare costs from all the above sources is specific to each man in the trial and the date of each cost item is also recorded. PCa-related costs could be identified using the PCa identifier available in the PMRR and, in the case of the CRHC data, using the ICD-10 code C61. We followed cost-analysis guidelines for the analysis of costs ([14–17]) and examined differences between the arms using two-sided

two-sample *t*-tests, with bootstrapping where appropriate to confirm the robustness of our results [18].

4. Results

Altogether, there were 31,867 men in the SA and 48,282 men in the CA (Fig. 1). Cost-related data were recorded in at least one of the registers used for 48,097 men in the CA (100%) and 31,753 men in the SA (100%). Cost records were not found for 198 men in the CA and for 119 men in the SA. These men may not have used hospital care or may not have been reimbursed for prescription medications during this follow-up period. No records from either register were found for one man in the SA who was diagnosed with, and subsequently died of, PCa. The frequencies of primary treatments were as follows: surgery (SA 26%, CA 19%), radiotherapy (SA 35%, CA 40%), endocrine treatment (SA 15%, CA 20%) and expectant management (SA 23%, CA 18%), with primary treatment missing for 1–2% of men in each arm. After 20 years of follow-up, healthcare costs were around 40% higher on average for men diagnosed with PCa, and around 70% higher on average for men who died from PCa, compared with all men in the trial (Table 1). The mean healthcare costs of the 80,149 men in the FinRSPC did not differ markedly between the arms, but a statistically significant difference, with lower costs in the CA, was observed when 'extreme

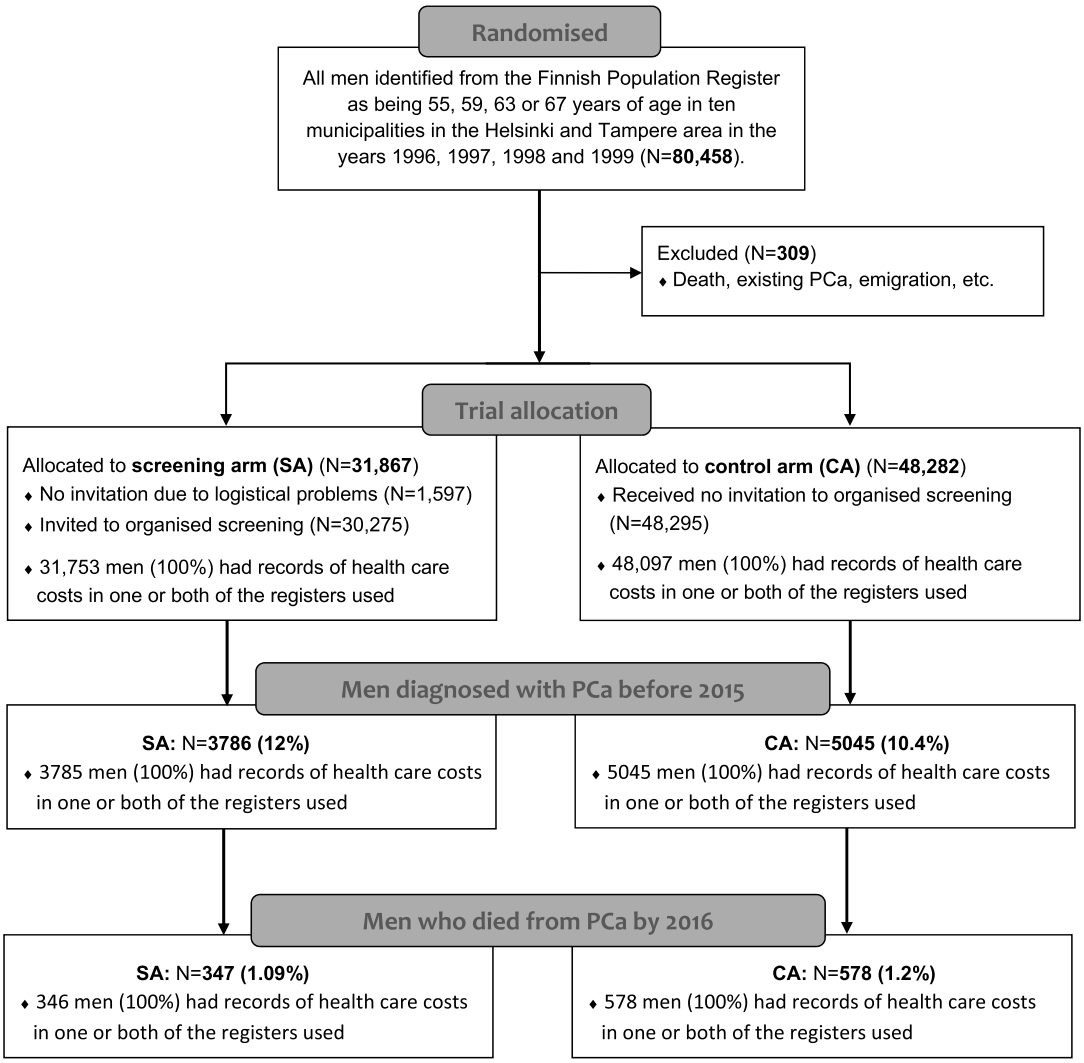

Fig. 1. Flowchart illustrating the groups for which register-based healthcare costs are estimated.

Table 1
Comparisons and statistical tests of the cost estimates.

Estimated register-based healthcare costs	Median in CA *(IQR in CA)*	Median in SA *(IQR in SA)*	Mean in CA	Mean in SA	Difference between means *(standard error)*	Two-sided *t*-test	Cohen's *d* effect size	Mean in SA, according to mode of detection
All-cause cost estimates for **all men in the trial**	24,900 € *(32,700 €)*	25,400 € *(33,000 €)*	**36,500** €	**36,300** €	**−200** € *(400 €)*	$p = 0.64$	<0.01	N.R.
All-cause cost estimates for **all men in the trial** (excluding 'extreme' observations†)	24,900 € *(32,700 €)*	25,400 € *(33,000 €)*	31,800 €	32,200 €	**400** € *(200 €)*	$\boldsymbol{p < 0.05}$	≈0.01	N.R.
All-cause cost estimates for **diagnosed men**	39,200 € *(42,500 €)*	38,800 € *(40,500 €)*	**51,900** €	**51,300** €	**−600** € *(1100 €)*	$p = 0.59$	≈0.01	**SD**: 50,300 € **CD**: 52,000 €
PCa-related cost estimates for **diagnosed men**	8800 € *(12,800 €)*	7700 € *(12,600 €)*	**15,300** €	**14,200** €	**−1100** € *(400 €)*	$\boldsymbol{p < 0.01}$	≈0.06	**SD**: 14,700 € **CD**: 13,800 €
PCa-related cost estimates for men with **low-risk**†† tumour at diagnosis	5900 € *(9700 €)*	5700 € (8900 €)	11,600 €	11,000 €	**−600** € *(600 €)*	$p = 0.35$	≈0.04	**SD**: 11,400 € **CD**: 10,200 €
PCa-related cost estimates for men with **intermediate-risk** tumour at diagnosis	8100 € *(9400 €)*	7800 € *(10,100 €)*	13,200 €	13,400 €	**200** € *(700 €)*	$p = 0.76$	≈-0.01	**SD**: 15,000 € **CD**: 12,500 €
PCa-related cost estimates for men with **high-risk** tumour at diagnosis	12,300 € *(17,700 €)*	11,800 € *(18,700 €)*	20,600 €	20,400 €	**−200** € *(1000 €)*	$p = 0.85$	<0.01	**SD**: 24,500 € **CD**: 18,400 €
PCa-related cost estimates for men with **metastatic** tumour at diagnosis	15,000 € *(25,100 €)*	15,600 € *(26,100 €)*	23,900 €	24,000 €	**100** € *(2000 €)*	$p = 0.94$	>−0.01	**SD**: 36,000 € **CD**: 21,900 €
PCa-related cost estimates for men with tumour information **missing** at diagnosis	6800 € *(7800 €)*	7400 € *(9200 €)*	8500 €	8500 €	**0** € *(800 €)*	$p = 0.95$	<0.01	**SD**: 6500 € **CD**: 8500 €
All-cause cost estimates for **men who have died from PCa**	47,600 € *(57,100 €)*	51,900 € *(53,400 €)*	**60,000** €	**65,100** €	**5100** € *(4600 €)*	$p = 0.27$	≈−0.08	**SD**: 71,000 € **CD**: 62,400 €
PCa-related cost estimates for **men who have died from PCa**	22,100 € *(35,600 €)*	23,600 € *(38,600 €)*	**31,400** €	**33,000** €	**1700** € *(2100 €)*	$p = 0.43$	≈−0.05	**SD**: 43,700 € **CD**: 27,900 €

Abbreviations: CA, control arm; SA, screening arm; IQR, interquartile range; N.R., Not relevant SD, PCa detected via the screening intervention; CD, clinically detected PCa.
† The cut-off point for extreme observations used was: 3rd quartile + (3*IQR).
†† We used the stage classification used by the European Randomized Study of Screening for Prostate Cancer (ERSPC) [1].

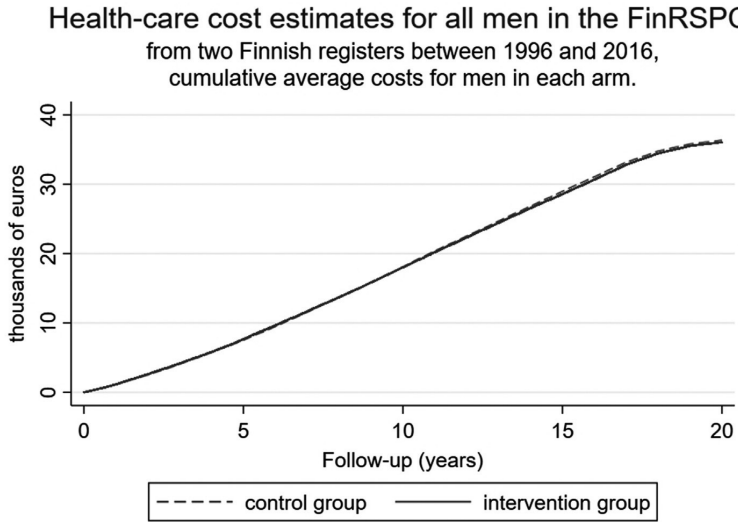

Fig. 2. All-cause cost estimates of healthcare and medications between 1996 and 2016 for all men in the trial (N = 80,149) by arm, cumulative average costs adjusted to the size of each trial arm.

observations' were excluded from our analysis (Table 1). However, the difference between the two arms in terms of cumulative all-cause costs appears to be small (Fig. 2). Further, all-cause healthcare costs for men diagnosed with PCa, adjusted for the number of men in each arm, were higher in the SA, with a steadily increasing differential (Fig. 3). Similarly, higher PCa-related cost estimates were seen for diagnosed men in the SA, with noticeable differences at follow-up years 1, 4 and 8 (corresponding to the screening rounds; Fig. 4). For healthcare costs of men who died from PCa during the follow-up, the difference increased after follow-up years 5 and 10 (corresponding to a year or more after the first and second screening rounds), with higher costs for the SA (Fig. 5). Similar findings were obtained when the graph for the men who eventually died from PCa was restricted to focus only on PCa-related average costs (Fig. 6). In Table 1, we also use the risk

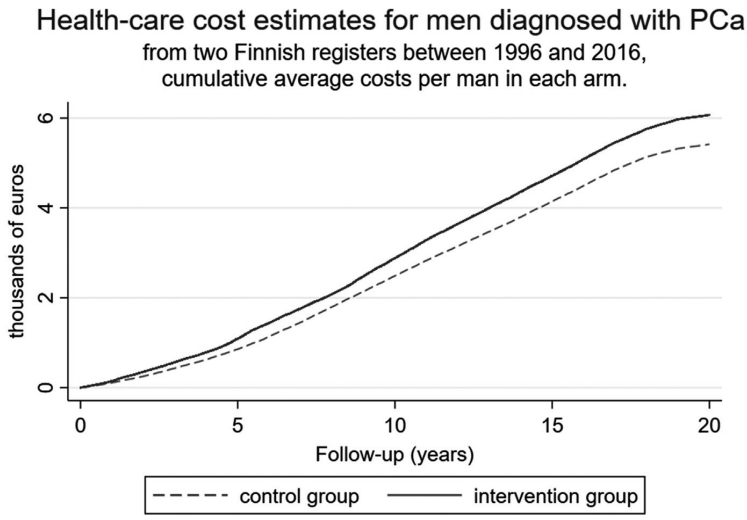

Fig. 3. All-cause cost estimates of healthcare and medications between 1996 and 2016 for men diagnosed with prostate cancer in the trial by arm, cumulative average costs adjusted to the size of each trial arm.

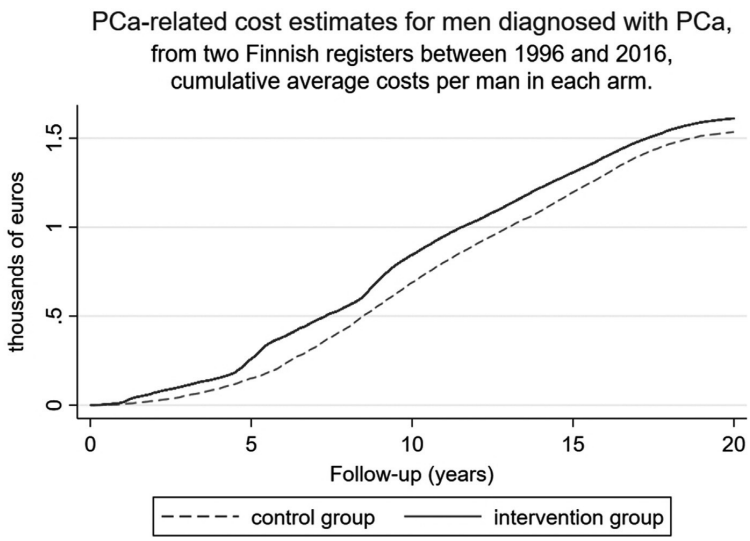

Fig. 4. Prostate cancer -related cost estimates of healthcare and medications between 1996 and 2016 for men diagnosed with prostate cancer by arm, cumulative average costs adjusted to the size of each trial arm.

Fig. 5. All-cause cost estimates of healthcare and medications between 1996 and 2016 for the men who have died from prostate cancer, cumulative average costs per man who died from prostate cancer in each trial arm.

classification utilised by the ERSPC to show how the PCa-related cost estimates for all diagnosed men vary with risk-stage at diagnosis. When comparing average PCa-related cost estimates for all diagnosed men, there is a statistically significant difference, with the SA incurring lower costs. However, this result is not observed for all the separate risk-stage subgroups. The mean cost estimates for PCa-related healthcare for men with low-risk, intermediate-risk, high-risk, or metastatic tumours at diagnosis and those for men with tumour information missing at diagnosis, although not statistically significant, are either higher for the CA, higher for the SA, or do not show differences in mean costs. As an adjunct to the ITS analysis, for the SA, we report the mean costs for both those prostate cancers detected due to screening (SD) and for clinically detected PCa (CD; Table 1). The effect sizes presented indicate that the comparison of PCa-related costs for men who died from

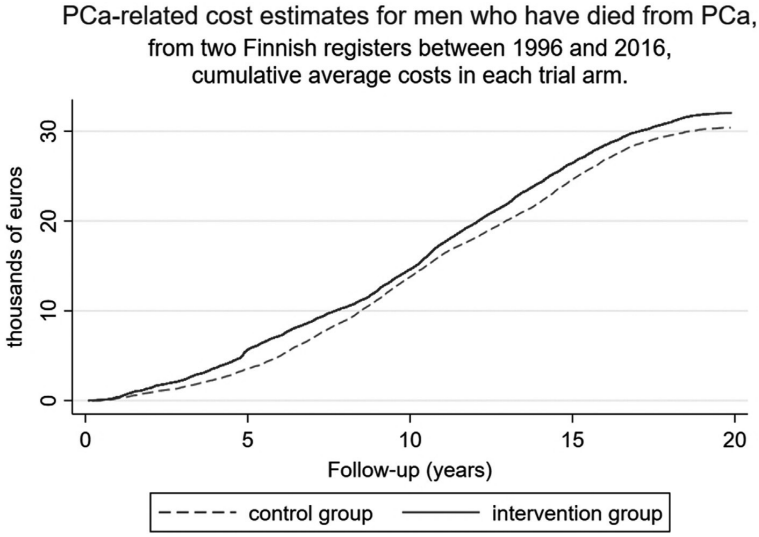

Fig. 6. Prostate cancer -related cost estimates of healthcare and medications between 1996 and 2016 for the men who have died from prostate cancer, cumulative average costs per man who died from prostate cancer in each trial arm.

PCa has the largest effect size (i.e., substantive significance, as indicated by the Cohen's *d* -measure), although this comparison of means did not achieve conventional levels of statistical significance (Table 1).

5. Discussion

For all men diagnosed with PCa, screening reduced mean PCa-related costs (by around 1100€, or less than 10% [see Table 1]). However, classification of PCa-related costs by risk-stage at diagnosis indicates that this result may be subject to a 'reversal paradox' [19], whereby this overall result may not faithfully represent the direction or size of each of the risk-stage subgroups. Further, for all diagnosed men, the reduction in mean all-cause healthcare costs was less than when focussing on PCa-related costs alone (i.e., around 700€, or around 1%). In addition, cumulative PCa-related healthcare cost estimates for these same diagnosed men, adjusted for the number of men in each trial arm, were slightly higher in the SA (around 100€, or less than 10% [see Fig. 4]).

Two important issues need to be considered when interpreting all our findings, first and foremost, the sample size required to show statistically significant results, second, opportunistic PSA testing (also known as contamination). First, the statistical power calculations for the FinRSPC are based on the primary outcome of disease-specific mortality. Therefore, the FinRSPC is not powered to evaluate all-cause costs, any more than it is suitably powered to detect differences in all-cause mortality, as only a minority of total mortality is directly related to PCa. In addition, the original sample-size estimates would likely need to be at least doubled to take into account the unexpected levels of contamination encountered in the trial [20], alternatively, the duration of follow-up would need to be extended beyond 20 years. Further, the comparisons of the relatively small subgroup of men who died from PCa (N = 925) have fairly low precision given the observed heterogeneity in costs between patients. Second, unorganised or non-systematic PSA-testing can dilute the observed effect of the mass-screening intervention. Instead of the comparator being a complete absence of screening, the comparator in the FinRSPC is less organised and less systematic screening. Widespread contamination likely dilutes PCa-mortality benefit, any differences in health-related quality of life, as well as the differences in costs observed here. Therefore, our results should be interpreted against the possibility of high levels of contamination in the CA, reflected for instance in the high cumulative incidence of T1c cancers (i.e., impalpable cancers detectable only by PSA testing [21]) in the CA, with a cumulative incidence of 4.5% in the CA and 6.1% in the SA [22].

It should also be noted that for men diagnosed with PCa, the overall costs were higher for the SA than for the CA, when adjusted for the size of each trial arm (Figs. 3 and 4), even though the mean healthcare costs for all men diagnosed with PCa were lower in the SA (Table 1). Importantly, these all-cause cumulative cost differentials could be explained by men with indolent disease being followed up clinically over extensive periods of time due to overdiagnosis. These cases could also involve some lead-time, increasing total costs. Similarly, the higher mean PCa-related costs in the SA men who died from

PCa could be due to some of these first 925 recorded PCa deaths including some of the more aggressive and rapidly progressing cancers. Precision medicine, with treatment tailored to the underlying molecular aberrations, holds promise for treating advanced PCa, but such interventions are currently at a largely experimental stage and not widely used. Targeted treatment has the potential to change the economic impact of screening, but currently it is impossible to predict whether early detection by screening will allow definitive treatment (with its potential for cost savings, increased life expectancy or improved quality of life), or merely delay disease progression (with its potential for increasing costs or decreased quality of life). In large part, any impact on costs will depend on both the differences in the time patients live with advanced PCa, as well as on the relative mortality, between the two trial arms.

One further interpretation of our risk-stage subgroup analysis and analysis of screen-detected versus clinically detected PCa (Table 1, and from other analyses not reported here due to restricted space) suggests men in the SA diagnosed via PSA-screening could have received more systematic care, or just more care in general, than those men in the CA. Analysis suggests that mean all-cause healthcare costs were lower for men with screen-detected PCa than for men in the screening arm overall, and lower for men with screen-detected low-risk tumour at diagnosis than for men in the screening arm overall. However, for all other diagnosed men, mean healthcare costs were higher amongst men with screen-detected prostate cancer than for men in the screening arm overall. Furthermore, the increasing differential observed in all-cause cost estimates for men diagnosed with PCa could also be explained by the screening intervention resulting in an increased awareness of health issues, or simply an increased supply of, or demand for, health services not directly related to PCa.

5.1. Strengths of the study

The novelty of our results is emphasised by the fact that a systematic literature search failed to identify studies reporting the analysis of real-world data on PCa-screening–related healthcare costs from any RCT, despite finding a number of studies on related topics (e.g. [23–31]).

Modelling studies, such as those found during our systematic search, offer estimates or forecasts of costs which are based on assumptions, which often, in turn, are based on modelled estimates of primary or secondary outcomes. On the other hand, our analysis describes the cost data recorded on the basis of observed outcomes. Such description will have relevance for economic evaluation using data from other ERSPC countries to the extent that, e.g., trial protocols are comparable. For this reason, despite only being from one participating centre in the ERSPC, our results may still be highly indicative of the relative difference in costs between the trial arms in other European countries. We were able to apply ITS analysis on a large and representative population over a 20-year period, with fairly comprehensive data on costs of hospital-care and prescription medication use and, hence, obtain accurate and potentially generalisable cost estimates. Further, our data-driven approach requires few assumptions concerning costs or outcomes for the men followed up in the FinRSPC over 20 years. Our data also capture most costs arising from the major disadvantage of PSA screening, i.e., the overdiagnosis of indolent PCa. This is evidenced by, e.g., the overall costs due to PCa being higher in the SA than in the CA (Figs. 3 and 4), despite men in the SA having lower PCa-related mean costs (Table 1). Although the precision of our results is adversely affected by the observed heterogeneity in costs, we have extensive observations from a publicly funded and centralised healthcare system, with highly comparable data over the study period.

5.2. Limitations of the study

First, our study covers a period which included major changes in PCa-treatment protocols. However, this limitation would be true of any pragmatic study in this field. The CRHC does not always provide sufficient information on procedures to provide precise details of all treatments for all periods; therefore, we reported the frequencies of primary treatments using the high-quality FinRSPC trial database. Secondly, we were unable to cover primary healthcare costs because, until recently (2014), no national primary healthcare registers or databases existed in Finland. Although data including primary-care costs were collected using questionnaires alongside the FinRSPC [10], that questionnaire data do not provide comparable data to the data used in this register-based study. However, the main responsibility for PCa management is with tertiary care, including the most expensive therapeutic procedures. A related limitation is the possibly limited applicability of our analysis outside the context of countries with mainly publicly funded healthcare. Third, consistent cost weights were not available for each year during our study period. Although we used the 2009 cost weights for all years, and these weights were not adjusted in any way, not accounting for health-sector inflation and not undertaking discounting, it seems unlikely that this would have a large negative impact on the policy relevance of our study. Any such adjustments to the cost estimates would likely affect the two arms equally [32], and our choice to round these estimates to the nearest 100€ is likely to negate any such adjustments in any case [33]. Perhaps most importantly, we have attempted to present costs in an appropriate manner for a policy-oriented readership. Unfortunately, no robust method seems to exist to extrapolate the observed cost estimates from

those men who have already died, to those who may die from PCa in the coming years [34]. A related limitation is that although potentially declining during the study period, overdiagnosis or overtreatment may still be one contributing factor in our results [35]. Despite such limitations, our estimates are expected to be indicative and representative of the main costs drivers in a publicly funded healthcare system, even though they do not represent the exact costs of all services used as a result of screening. Further, the analysis presented here does not provide a definitive assessment of the impact of costs on PSA-based mass screening. The full impact of screening on healthcare costs will only be clear after all the men are deceased, and the FinRSPC cohort is relatively immature in this respect (63% of men are still alive). Of all men, 1.2% had died from PCa during the 20-year follow-up period, whereas the expectation is that PCa mortality will eventually reach over 2% in this population [2].

6. Conclusion

No major cost impacts due to screening were apparent, but after 20 years of follow-up, the FinRSPC trial shows that for all diagnosed men, mean PCa-related costs were lower in the screening arm. However, in addition, mean healthcare costs for the men who died from PCa appear to be substantively higher in the screening arm. These estimates of differences in mean healthcare costs should be interpreted in the light of low statistical power, the effects of PSA-contamination within the trial, and with the knowledge that these estimates of average costs may be impacted by extreme observations and cover up differences between risk subgroups. In conclusion, the 20-year follow-up of this large cohort is too short to give definitive evidence about the healthcare costs of PSA screening. Longer term follow-up will be required to be better informed about the costs of, or savings from, introducing PSA-based mass screening.

Conflict of interest statement

Kimmo Taari has taken part in a congress with support from Astellas and received research funding from Medivation, Astellas, Pfizer and Orion. Teuvo Tammela has acted as a consultant for Orion Pharma, Bayer AG and Ferring and received research funding from Medivation, Pfizer and Lidds Ab. All other authors have nothing to disclose.

Roles of the funding sources

The funding allowed collection of the data over a long period and allowed the corresponding author to work with the data and postpone analyses until more data were available, rather than being pressured into early publication. There was no role of for-profit healthcare companies in the writing of the manuscript.

Acknowledgements

This study was financially supported by the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (grants 9E089, 9F100, 9G096, 9H099, 9L085, 9N064 and 9R002) and by the Yrjö Jahnsson Foundation (grants 6213 and 6572).

The authors would like to express thanks to all those who have helped in collating the data over the past twenty years, but especially to Eeva Kangasniemi and Hilkka Lamminsivu from Pirkanmaa Hospital District, to Kristiina Tyrkkö from the Social Insurance Institution of Finland, as well as to Liisa Määttänen from the Mass Screening Registry. The corresponding author would also like to thank Jani Raitanen and Pasi Aronen for numerous helpful discussions about data manipulation and matters statistical.

Appendix

The FinRSPC trial database contains information about all PCa diagnoses before 2015 obtained from the nationwide Finnish Cancer Registry, (FCR). The FCR has been shown to have comprehensive coverage of all solid cancers diagnosed in Finland [12], but these diagnoses data from the FCR were also confirmed from medical records as part of the FinRSPC. This trial database also includes data from Statistics Finland's Causes-of-death statistics (available from: https://www.stat.fi/til/ksyyt/index_en.html), which has, since the start of the FinRSPC in 1996, applied the 10th revision of the International Classification of Diseases (ICD-10). For a sample of men from the FinRSPC, the official causes of death were reviewed by an independent expert review panel and found to be in close agreement (overall agreement 98%, $\kappa = 0.95$) [13]. For the period of the study (1996–2016), the two main registers containing information on healthcare utilisation and costs were the CRHC and the PMRR, i.e., the Care Register for Health Care ("Hilmo" in Finnish, produced by THL) and the Prescription-Medicine Reimbursement Register ("Lääkeostotiedot" in Finnish, produced by the Social Insurance Institution of Finland [Kela]). Since 2010, the CRHC has consistently included records of both outpatient and inpatient visits to both secondary and tertiary health care. The interested reader can refer to https://www.thl.fi/en/web/thlfi-en/statistics/information-on-statistics/register-descriptions/care-register-for-health-care for further details on the CRHC. Before 2010, only inpatient records were available from the CRHC, so for this earlier period, we used all available outpatient records

collated by the two hospital administrations in the districts of Pirkanmaa and Uusimaa. On the advice of the producers of the CRHC, the 2009 Nordic diagnosis-related groups (NordDRG) classification system is used here for inpatient care episodes before 2010. For episodes in 2010 and later, the 2015 NordDRG classifiers are used. Whenever it is possible to calculate a NordDRG-based cost for both inpatient and outpatient costs, the 2009 NordDRG costing weights are applied. When this is not possible (in a few instances and largely before 2010), we use the municipal billing records from the administrative databases in the hospital districts of Pirkanmaa and Uusimaa to estimate the costs. The PMRR contains the exact costs of prescription medications sold through retail pharmacies but does not include information on prescription medications supplied by hospital pharmacies. Another restriction of the nationwide PMRR is that information limited to only those prescription medications reimbursed under the Health Insurance Act at any point in time. The interested reader can refer to http://www.kela.fi/web/en/inclusion-of-medicines for further information on the reimbursement status of medicines.

References

[1] Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the european randomised study of screening for prostate cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384(9959):2027–35.

[2] Streetly A, Elhers L. Population screening and public health. In: Detels R, Gulliford M, Karim QA, editors. Oxford textbook of global public health. Oxford University Press; 2015. p. 1507–22.

[3] McKeown T, Butterfield WJH, Cochrane AL, Cohen RHL, Knox EG, Lowe CR, et al. Screening in medical care: reviewing the evidence. Nuffield Provincial Hospital Trust; 1968.

[4] Russell LB. Educated guesses: making policy about medical screening tests. Berkley: University of California Press & Milbank Memorial Fund; 1994.

[5] Sackett DL. The arrogance of preventive medicine. CMAJ (Can Med Assoc J) 2002;167(4):363–4.

[6] Glick H, Doshi J, Sonnad S, Polsky D. Economic evaluation in clinical trials. Oxford: Oxford University Press; 2014.

[7] Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Malik F, et al. Transferability of economic evaluations across jurisdictions: ISPOR good research practices task force report. Value Health 2009;12(4):409–18.

[8] Hakama M, Aro J, Auvinen A, Juusela H, Määttänen L, Stenman UH, et al. Randomized screening trial for prostate cancer in Finland. Eur Urol 2001;39(Suppl 4):32.

[9] Finne P, Stenman UH, Määttänen L, Mäkinen T, Tammela TL, Martikainen P, et al. The Finnish trial of prostate cancer screening: where are we now? BJU Int 2003;92(Suppl. 2):22–6.

[10] Booth N, Rissanen P, Tammela TLJ, Määttänen L, Taari K, Auvinen A. Health-related quality of life in the Finnish trial of screening for prostate cancer. Eur Urol 2014;65(1):39–47.

[11] Anttila A, Lönnberg S, Ponti A, Suonio E, Villain P, Coebergh JW, et al. Towards better implementation of cancer screening in Europe through improved monitoring and evaluation and greater engagement of cancer registries. EJC (Eur J Cancer) 2015;51(2):241–51.

[12] Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. Acta Oncol 1994;33(4):365–9.

[13] Mäkinen T, Karhunen P, Aro J, Lahtela J, Määttänen L, Auvinen A. Assessment of causes of death in a prostate cancer screening trial. Int J Canc 2008;122(2):413–7.

[14] Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS)—explanation and elaboration: a report of the ISPOR health economic evaluation publication guidelines good reporting practices task force. Value Health 2013;16(2):231–50.

[15] European Network for Health Technology Assessment [EUnetHTA]. Methods for health economic evaluations—a guideline based on current practices in Europe. 2015. Available from: http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/2015-04-29-ECO-GL_Final%20version_0.pdf.

[16] Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE, Ganiats TG, editors. Cost-effectiveness in health and medicine. 2nd ed. Oxford: Oxford University Press; 2016.

[17] Sullivan GM, Feinn R. Using effect size-or Why the P Value is not enough. J Grad Med Educ 2012;4(3):279–82.

[18] Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000;320(7243):1197–200.

[19] Tu Y-K, Gunnell D, Gilthorpe MS. Simpson's Paradox, Lord's Paradox, and Suppression Effects are the same phenomenon – the reversal paradox. Emerg Themes Epidemiol 2008;5:2.

[20] Torgerson DJ. Contamination in trials: is cluster randomisation the answer? BMJ 2001;322(7282):355–7.

[21] Sobin LH, Wittekind C, Gospodarowicz MK. TNM classification of malignant tumours. Oxford: Wiley-Blackwell; 2010.

[22] Kilpeläinen TP, Pogodin-Hannolainen D, Kemppainen K, Talala K, Raitanen J, Taari K, et al. Estimate of opportunistic prostate specific antigen testing in the Finnish randomized study of screening for prostate cancer. J Urol 2017;198(1):50–7.

[23] Roth JA, Gulati R, Gore JL, Cooperberg MR, Etzioni R. Economic analysis of prostate-specific antigen screening and selective treatment strategies. JAMA Oncol 2016;2(7):890–8.

[24] Shteynshlyuger A, Andriole GL. Cost-effectiveness of prostate specific antigen screening in the United States: extrapolating from the european study of screening for prostate cancer. J Urol 2011;185(3):828–32.

[25] Pataky R, Gulati R, Etzioni R, Black P, Chi KN, Coldman AJ, et al. Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada. Int J Cancer 2014;135(4):939–47.

[26] Shin S, Kim YH, Hwang JS, Lee YJ, Lee SM, Ahn J. Economic evaluation of prostate cancer screening test as a national cancer screening program in South Korea. Asian Pac J Cancer Prev APJCP 2014;15(8):3383–9.

[27] Torvinen S, Färkkilä N, Roine RP, Sintonen H, Saarto T, Taari K. Costs in different states of prostate cancer. Acta Oncol 2016;55(1):30–7.

[28] Hall PS, Hamilton P, Hulme CT, Meads DM, Jones H, Newsham A, et al. Costs of cancer care for use in economic evaluation: a UK analysis of patient-level routine health system data. Br J Cancer 2015;112(5):948–56.

[29] Rasmussen JF, Siersma V, Pedersen JH, Heleno B, Saghir Z, Brodersen J. Healthcare costs in the Danish randomised controlled lung cancer CT-screening trial: a registry study. Lung Canc 2014;83(3):347–55.

[30] Mäklin S, Hakama M, Rissanen P, Malila N. Use of hospital resources in the Finnish colorectal cancer screening programme: a randomised health services study. BMJ Open Gastroenterol 2015;2(1):e000063.

[31] Heijnsdijk EAM, de Carvalho TM, Auvinen A, Zappa M, Nelen V, Kwiatkowski M, et al. Cost-effectiveness of prostate

cancer screening: a simulation study based on ERSPC data. J Natl Cancer Inst 2015;107(1).

[32] Mauskopf JA. Budget-impact analysis. In: Culyer AJ, editor. Encyclopedia of health economics. Burlington: Elsevier Science; 2014.

[33] The European network for Health Technology Assessment (EUnetHTA) project. HTA Core Model: version 3.0. Available from: http://www.corehta.info/model/HTACoreModel3.0.pdf.

[34] Bojke L, Manca A, Asaria M, Mahon R, Ren S, Palmer S. How to appropriately extrapolate costs and utilities in cost-effectiveness analysis. Pharmacoeconomics 2017;35(8):767–76.

[35] Roobol MJ, Schröder FH. The rate of overdiagnosis inextricably linked to prostate-specific antigen -based screening for prostate cancer can Be quantified in several ways, but what is the practicable message? Eur Urol 2014;65(6):1056–7.

PUBLICATION III

On value frameworks and opportunity costs in health technology assessment

Neill Booth

International Journal of Technology Assessment in Health Care, 35(5), 367-372

International Journal of Technology Assessment in Health Care

cambridge.org/thc

Commentary

Cite this article: Booth N (2019). On value frameworks and opportunity costs in health technology assessment. *International Journal of Technology Assessment in Health Care* 1–6. https://doi.org/10.1017/S0266462319000643

Received: 18 March 2019
Revised: 6 August 2019
Accepted: 15 August 2019

Key words:
Costs and cost analyses; Priorities, Health; Health; Technology assessment, Health

Author for correspondence:
Neill Booth, E-mail: neill.booth@uta.fi

CAMBRIDGE UNIVERSITY PRESS

On value frameworks and opportunity costs in health technology assessment

Neill Booth

Faculty of Social Sciences, Tampere University, Tampere, Finland

Abstract

Objectives. Proceeding from a basic concept underpinning economic evaluation, opportunity cost, this study aims to explain how different approaches to economics diverge quite dramatically in their ideas of what constitutes appropriate valuation, both in principle and practice. Because the concept of opportunity cost does not inherently specify how valuation should be undertaken or specify how appropriate any economic value framework (EVF) might be, the three main economics-based approaches to providing evidence about value for health technology assessment are described.

Methods. This paper describes how the three main EVFs—namely, the extra-welfarist, welfarist, and classical—are most typically understood, applied, and promoted. It then provides clarification and assessment of related concepts and terminology.

Results. Although EVFs differ, certain underlying characteristics of valuation were identified as fundamental to all approaches to economic evaluation in practice. The study also suggests that some of the rhetoric and terms employed in relation to the extra-welfarist approach are not wholly justified and, further, that only the welfarist approach ensures adherence to welfare-economic principles. Finally, deliberative analysis, especially when connected with a classical economic approach, can serve as a useful supplement to other analytical approaches.

Conclusions. All three approaches to economic evaluation have something to offer assessment processes, but they all display limitations too. Therefore, the author concludes that the language of economic evaluation should be used with sufficient humility to prevent overselling of EVFs, especially with regard to the qualities of evidence they provide for priority setting processes.

In a recent commentary piece in this journal (1), Professor Culyer usefully highlighted many of the issues in economics surrounding costs and context in health-economic evaluation for health-technology assessment (HTA). Although making appropriate reference to both health economics and economics in general, Culyer neglected to mention that economics for HTA can encompass more than the extra-welfarist approach and that other economic value frameworks (EVFs) exist. Although his commentary helps to demystify the topic, still greater clarity and humility with regard to "economic" perspectives on valuation could contribute to improved HTA processes. Indeed, assessing the quality and relevance of EVF outputs as information for priority setting processes may become easier once the fundamental assumptions and value judgments related to EVFs are clarified.

This paper highlights two main alternatives to extra-welfarist economic value frameworks (EWEVFs)—the welfarist (WEVF) and the classical (CEVF)—and it describes how both can inform HTA decision making processes. Each of the three economic approaches here depends on particular sets of premises (in essence, "political" judgments) as to which sorts of value count and the extent to which those dimensions of value are covered. Hence, as Culyer and Jönsson note (2, p. 2), these can be seen as vital for correctly judging the applicability or relevance of any given EVF.

Theoretical Understanding of Opportunity Cost

This paper supplements earlier studies by clarifying several factors related to economic evaluation for HTA. Conceptual clarity is especially important both when defining opportunity costs and when actually carrying out any corresponding economic evaluation, on account of the implicit or explicit assumptions made, the limitations and uncertainties surrounding the measurement instruments, and the challenges involved in estimating any form of "economic" efficiency. A clear, transparent approach is important also with regard to terminology: as Williams argued several decades ago, the role of economic evaluation in setting priorities for health technologies is easily oversold (3), and the relevance of this has been reaffirmed many times since (4;5). Another important reason to strive for clarity lies in a shift witnessed in economic evaluation away from more welfarist views (6, p. 64) and toward more narrowly focused extra-welfarist EVFs (7). Although the Culyer piece offers a textbook parable related to opportunity cost, it bears remembering that economists have utilized the concept at least as far

back as Adam Smith's day (8, Book I, Chapter VI, p. 1). The term "opportunity cost" itself was coined by Green, with the thrust of his definition already involving "the opportunities foregone in accepting a certain line of action" (9). Differences between schools of economic thought notwithstanding, Green's definition seems to have been reinforced—by, among others, both Alchian (10) and Buchanan, with the latter stating that "opportunity cost is the evaluation placed on the most highly valued of the rejected alternatives or opportunities" (11). Though there is fairly widespread agreement that economic evaluation is intended to *inform* HTA decision making processes, how this principle gets applied in practical analysis of opportunity costs will reflect both the policy problems facing decision makers and the research questions involved, along with the specific EVF chosen (1;12).

At the conceptual level, identifying opportunity costs entails a two-part approach: first, the value of the "new" technology at issue is estimated or defined; then, the estimate obtained is compared with the value placed on the class of all "practicable" alternative technologies, however specified. The first of the two evaluative components assigns a value to the given health technology relative to at least one other way of serving the same group. This valuation addresses not only the estimated additional resource requirements of the new technology, but also takes into account its effectiveness; that is, this first valuation reports or estimates a value for at least one of the outcomes produced by the health technology. The second component places a value on what would have to be forgone for to supply the resources needed for the chosen technology. The objective of any reputable economic evaluation is therefore to provide evidence on whether the technology's economic value (ascertained in the first component) outweighs the economic value of what is foregone (ascertained in the second component). The likely utility of economic evaluation for decision making purposes is markedly lower when either of the two evaluative components lacks plausibility. Accordingly, this paper focuses on clarifying the nature of economic evaluations' information inputs to priority setting processes. From this perspective, it outlines the orientation of three EVFs, which, to varying extents, can address policy problems and identify different forms of opportunity cost (1). The aim is a critical review of economists' attempts to adopt and operationalize these concepts, bundled as they are with particular aspirations, conditions, and premises.

Concepts of Opportunity Cost in Practice

There are three main "economics"-based approaches to determining whether a given technology's economic value exceeds the value of any action forgone. Each type of EVF—the extra-welfarist, the welfarist, or the classical—imposes its own boundaries on how the valuation is undertaken. For each of the two components described above, the frameworks typically identify (or tacitly accepts) their own sources of "value" and/or metrics thereof. These differences between EVFs stem predominantly from *what* is deemed to be of value, though EVFs also diverge in *how* the valuation is conducted.

For a backdrop to examination of differences between EVFs, it is useful to outline the scope of investigations that are possible as part of the economic evaluation of health-care technologies. There are at least five distinct levels at which concepts of opportunity cost can be considered: (i) choices from among particular portfolios of public expenditure (13); (ii) choices from among the technology portfolios that constitute the basket of publicly provided services (7); (iii) choices between treatments within the limits set for total disease-specific expenditure (14); (iv) choices between mutually exclusive treatments (15); and (v) estimates of what may be forgone through using a specific input to the production process, or "resource opportunity cost" (16). The focus here is on level (ii), because the portfolio-of-technology level represents the most prevalent scope adopted by economic evaluations aimed at informing processes of health-care resource allocation (17).

Differences between EVFs

The objective for extra-welfarist approaches is often characterized as being to "maximize health" (18), where the matter of how "health" is defined can be considered very important because of proxying; typically in EWEVFs, rather than "health" *per se* being maximized, only an indicator of health is maximized. Under EWEVFs, "health" usually refers to the amalgam of (i) an indicator reflecting some dimensions of perceived health status with (ii) "health-state valuations" connected with that indicator (19). Both many of the indicators, and many of the valuations thereof, are typically engineered by health economists themselves. Although extra-welfarist approaches do not dictate a given maximand, most EWEVF applications center on maximizing a *combination* of precisely this sort of "social valuation" of states of health *with* estimates of length-of-life impacts, normally operationalized in the form of quality adjusted life-years (QALYs). Under EWEVFs, the first evaluative component's output, typically a cost-per-QALY estimate, is compared with the second "output," which represents "opportunity cost" (an estimated mean cost per unit of health forgone through diverting resources from other activities). Thus, in principle, EWEVFs address whether total "health" will increase if the new technology is introduced, but do so with an implicit assumption that both the new technology and the activities from which resources are diverted are, as economic theory suggests, perfectly divisible with constant returns to scale. However, as noted by Drummond (6) and illustrated by Birch and Donaldson (20), ascertaining the new technology's impact on efficiency (net impact on health) in a theoretically well-grounded manner requires avoiding such strict assumptions, which demands a mathematical-programing approach.

The aim with welfarist approaches to economic evaluation is to maximize "welfare," where analysis is undertaken to identify the improvements in the aggregate welfare of individuals (21). Valuation using WEVFs is based on the utility individuals gain from how the available resources are used, inclusive of any welfare impacts arising from the way commodities or outcomes are distributed within the population in connection with different uses of resources (22). "Social welfare" or "well-being" can be defined in terms of total net willingness to pay (WTP) (23), with contingent valuation methods constituting the main source of valuations in WEVFs (24). In more general terms, WEVF-based analysis compares the additional well-being produced by the new technology with that forgone through diversion of the required resources from elsewhere to support the new technology.

Finally, in classical approaches to economic evaluation, one of the central objectives is to supplement EWEVFs and WEVFs by accounting for preferences or values that are ascertainable only via deliberative methods. The label "classical" refers to the long history of valuation in economics before such developments as the marginal revolution (25). With CEVFs, the goal is to identify and assess, rather than to define and maximize, "health" or

"well-being." That is, in place of a formalized maximand, the targets in a classical approach (26, p. 136) might involve satisficing (27) or sufficiency (28), in addition to interpreting, for example, some EWEVF- or WEVF-derived indicator of "economic" efficiency. Often, CEVFs operate with other, non-quantitative information too, and typically encompass deliberation (29). Perhaps their most important element is an attempt to avoid being constrained to focus on formal economic efficiency, that is, on the type of neo-classical economic efficiency which is the result of quantitative or mathematical analysis.

EVFs, Opportunity Cost, and the Two Components of Valuation

As the name "economic value framework" suggests, each EVF has its own approach to valuation embedded within it. Under EWEVFs, one frequent approach to judging what is forgone is to assume, both in principle and practice, that it is possible to quantify an opportunity cost and that this quantity is invariant to the size of the program being evaluated, that is, that there can be a fixed "cost per QALY" (30). However, this is inconsistent with the economic notion of resource scarcity and the general finding that the marginal utility of a good or service decreases as consumption increases. When EVFs employ comparison to some fixed monetary valuation of opportunity cost, they tend to ignore factors such as the potential budgetary impact of the intervention and the "lumpiness" of health technologies (31;32).

Although all three EVFs entail estimating cost and effect differences for a new technology relative to a comparator, the discussion above should render it clear that there may be little deeper commonality in how EVFs assign value to alternative health technologies that might be displaced. The onus is generally on the user of the research to identify the possible implications of the chosen value system for the decision making process it is purported to serve (33). The discussion below attempts to make the relevant implications clearer for each of the three main EVFs.

Valuation and Opportunity Cost in EWEVFs

Under EWEVFs, the first evaluative component in defining opportunity cost is generally based on cost-effectiveness analysis, which yields an estimate of the mean cost-per-unit health benefit produced by the chosen intervention—that is, an incremental cost-effectiveness ratio (ICER). In EWEVFs, this ratio, an estimate of the inverse of the mean rate of return on the additional investment required to fund the technology, is typically employed in an economic-efficiency metric entailing comparison with some predetermined benchmark ICER, that is, some cost-effectiveness-ratio threshold (CERT) (34). The latter is usually exogenous to the study at hand. Only rarely under EWEVFs do the activities displaced by the additional investment of resources in the technology get identified, or be valued, on a case-by-case basis. Although some CERTs involve estimates from econometric analysis of possible relationships between current resource use and health-related outputs (35;36), they may also simply represent an arbitrary figure or diktat (37). Indeed, CERTs will generally fail to fully reflect the actual displacement resulting from the technology's adoption (38). Many researchers continue to propose CERTs, of various types, despite evidence suggesting that thresholds are merely an economic abstraction and that a single appropriate CERT is likely to remain elusive in most contexts (39).

WEVF-Related Valuation and Opportunity Cost

Under WEVFs, analysis focuses on individuals' preferences and technologies are evaluated for their impacts on "well-being" (20). In some of these frameworks, the two evaluative components are brought together in a single model for analysis of portfolio choice through mathematical optimization. By incorporating resource constraints into the model explicitly, thereby focusing attention on the well-being generated from the entire resource budget as opposed to a single program's share of that budget, the approach addresses opportunity cost considerations directly without requiring the separate valuation of the foregone alternatives that is typical under EWEVFs (40). Hence, the emphasis in WEVFs is on comparing across the well-being generated by various combinations (or portfolios) of "health technologies" that the available resources can sustain, and on determining which combinations could improve "welfare." In addition, the approach can accommodate any other concrete constraints on preferences, in line with policy considerations related to equity, need, and so on. (40). It is also important to note here that, in practice, WEVF utilizes WTP estimates which typically rely on methods such as contingent valuation to compare WTP between the new technology in aggregate and whatever must be forgone (41).

Valuation and Opportunity Cost in Classical Economic Approaches

CEVFs can be viewed as a reaction to various limitations of EWEVFs and WEVFs in practice, especially as the latter are designed to "maximize" via an objective function of one type or another. CEVFs represent an alternative approach, one that need not focus on a single maximand (as EWEVFs typically do) or on a single source of preferences (as is typical under WEVFs, the source being individuals) yet CEVFs can still be in line with conventional interpretations of opportunity cost (5).

How CEVF Approaches can Help in HTA

In light of the above, CEVFs are proposed as an alternative that affords wider scope than either "health maximization" under EWEVFs or "maximization of economic welfare" under WEVFs, as they allow for qualitative use of preferences from groups of individuals, or directly from other stakeholders. Rather than rejecting use of the other EVFs, the CEVF approach supplements them with further information or deliberative analysis, such as incorporating community values (42) canvassed through various evidence-gathering processes (43–45).

A CEVF approach can help inform HTA in three main ways. First, CEVFs can add information to evidence provided by EWEVF and WEVF approaches on the relative efficiency with which "health" and "welfare" are produced, respectively. Although WEVFs may include strong evidence about budget or resource impacts, additional, related information (with either a short or a long time horizon) can still be produced or utilized within a CEVF (46). Second, CEVFs can identify any qualifications or caveats to the EWEVF or WEVF findings, aiming to ensure that the information they provide is interpreted correctly, through an appropriate appraisal of their quality. Although such appraisal is already addressed by many existing HTA processes, it could have greater value due to being integral to a CEVF approach, in line with an iterative, classical vision of valuation (26). The third main advantage would be that CEVFs can provide fuller awareness of

the nature of the research question and its connection with the policy problem, as well as of the types and levels of uncertainty and relevance carried by information from other EVFs (47;48). One major contribution that CEVFs can make to HTA processes is to force more clarity into the terminology surrounding EVFs. This point will be returned to below.

CEVFs allow inclusion of dimensions of value that might not be measurable in the commensurate units "required" by EWEVFs or WEVFs (49). Because they can take into account informal analysis during an iterative process of deliberation, CEVFs could prove highly relevant for decision makers (50). This might involve, for instance, (a) confirming, doubting, or disproving the suitability of standard health-economic outcome metrics for the technology in question, partly through questioning the assumptions underlying information outputs from other EVFs, and (b) establishing additional objectives or outcome measurements relevant for the technology in question (51, p. 149). For item (a), deliberative analysis may assist in identifying any need to supplement other EVFs, because it is probable that no single overriding "efficiency" principle meets all the desiderata for allocation, and there may be good reasons to consider multiple prioritization principles (29). For instance, some opportunity costs may not be quantifiable (52) and might lend themselves only to deliberation, as in the case of rights-based deontological or paternalistic considerations (53). In addition, with regard to item (b), for some technologies there may be little pertinent quantitative information available from formal analysis, and stakeholders may hold diverse, conflicting views (54). The appraisal process may embody a range of considerations that might not all be well-defined prior to, or even during, economic evaluation. There are numerous situations in which deliberative analysis via CEVFs may provide a useful extension that improves on purely formal analysis, and a variety of evidentiary inputs may be used, as necessary, on a case-by-case basis (1;55).

In general, although analytic endeavors within EWEVFs or WEVFs can reveal some of the implications of particular choices (33), CEVFs may add a platform that stimulates discussion of more communitarian values (e.g., (56;57)). With CEVFs, the aim is what some have called "higher-level efficiency," rather than efficiency in the more neo-classical sense found in the more formal approaches of EWHEE and WHEE (58, p. 125).

Discussion

Each mode of economic thinking outlined in this paper can offer useful information for priority setting processes, even though each EVF involves its own particular aims, assumptions, and value judgments. Whichever EVF is applied, evaluating opportunity cost requires *some* valuation of what is given up (59); hence, the aim here is not to denigrate or promote any particular mode of economic evaluation but to promote solid awareness of the information that each can provide. In all cases, it should be acknowledged that economic approaches to assessing opportunity costs are information-intensive in their input requirements and that their use often suffers from a lack of appropriate information (60), especially as pathways to health are often quite complex (61). One should also bear in mind that any method which gives consistent or accountable answers in a systematic manner is unlikely to yield truly comprehensive evaluation (62). There are many circumstances wherein measurements fail to cover relevant aspects of the changes in "states of health" (32;63) or do not capture changes in capabilities or in patient-reported experiences, not to mention the fact that "social valuations" of such changes in the health status do not fully capture society's values (5). On account of the measurement issues surrounding WTP, there may be many situations in which no valid and reliable methods of operationalizing WEVFs exist (41;64).

Problems with the EVF Lexicon

Although choice processes for allocating health-care resources should lead to transparent mechanisms for valuation of the various options and their opportunity costs (65, p. 138), terminology can make economic evaluation more opaque. This is evident from the declining use of terminology relating to intangibles and incommensurability, which could be seen as arrogant in a subdiscipline that often preaches humility. On account of space restrictions, the discussion here focuses on the terms "cost," "threshold," "decision rule," and "value for money."

"Cost" has multiple meanings in both lay and specialist use, as Culyer noted when deeming it naïve to employ the term "cost" for undesirable attributes (1). An alternative interpretation to that offered by Culyer is to take the undesirable attributes of an intervention as also representing *a* cost. Of course, at the level of valuing what may be forgone through using a specific input to the production process, or "resource opportunity cost," that is, at the level of *building* the pool from Alchian's and Culyer's examples, then "undesirable attributes" should not be referred to as costs. On the other hand, the use of the term "cost" for an undesirable attribute, a harm, or a negative benefit, could legitimately be used to refer to its part in an estimate of higher-level opportunity cost, that is, when assessing the value of the pool *per se*. Indeed, at the portfolio-of-technology level, such undesirable attributes can be seen as an *essential component* of any EVF. Undesirable attributes are important when forming a valuation; Alchian expresses it thus: "The decision maker must choose among events that are amalgams of goods and bads" (10). Therefore, in addition to the *things* forgone, such as the financial costs and the resources tied up, other aspects of the *value* forgone, the "costs" in terms of harms to health will also have a legitimate place in economic evaluations' definitions of (opportunity) costs (66). In practice, economic evaluations do typically include undesirable attributes in their analysis; for instance, EWEVFs do tend to utilize something akin to Alchian's amalgam approach when they promote a metric expressing the estimated cost divided by the estimated incremental overall population-"health impact." For the purposes of HTA, it seems reasonable to suggest that any sound economic evaluation involves taking both pros and cons into account: focusing on both the undesirable and the desirable attributes of technology, in line with the foundations of technology assessment (67). Although, obviously, pain and suffering need not involve resources *per se*, the principle of opportunity cost encompasses the benefit forgone, so any robust measurement of higher-level opportunity cost should also take the "cost," in terms of related pain and suffering, into account.

Some extra-welfarist economists and even some HTA practitioners take the perspective that "thresholds" can and should be quantified. However, economizing in line with these assumptions may be less intuitive for others involved in prioritization processes and seem rather perfunctory with respect to "societal values" (68;69). As is noted above, defining opportunity cost as a single threshold estimate can be seen as a typical economic abstraction. Although economic evaluation must always operate at some level of abstraction in practice, the fairy tale of a single threshold

(CERT), or threshold range, can be regarded as unhelpful. As no such one-size-fits-all threshold exists in reality, even within a well-bounded single jurisdiction, employing the term "threshold" seems to oversell EWEVFs. The problematic terminology is compounded by the use of connected phrasings such as "decision rules" and "value for money." For instance, the real-world applicability of so-called decision rules of EWEVFs is crucially reliant on the framework's inherent value judgments and assumptions. Indeed, these "rules" are typically valid only within the confines of the EWEVF in question, and there is a danger that the term "decision rules" could be construed to carry a similar meaning beyond this arcane hypothetical setting. Furthermore, claims of ICERs revealing "value for money" seem quite arrogant, in that EWEVFs often offer only a highly abstracted indicator of value. Although the concise term "value for money" may be much easier to sell to HTA decision makers than, for example, "estimated mean valuation of estimated change in mean health status divided by the estimated change in mean health-care costs," the former loses too much in precision; it seems much less honest. Because loose language could result in dire consequences of economic evaluation being oversold to the HTA community, it should be avoided at all costs.

Conclusions

Rather than economists holding a uniform, all-encompassing view, there are three main approaches to economic thinking for HTA, accompanied by a multitude of ways to implement each of these. Instead of a single notion of economics embodied by one EVF, the study found EWEVFs, WEVFs, and CEVFs, each with the corresponding problems and potential. Therefore, all approaches to economic evaluation should be checked for quality and relevance before being used to inform prioritization processes. Applying more precise vocabulary, coupled with greater understanding of the limits to analysis of any kind, should help decision makers engage in appropriate deliberation and interpretation in their HTA endeavors. The ways in which notions of opportunity cost are translated into practice and interpreted are likely to have great importance, not only for priority setting but also for the long-term health and sustainability of health-care systems.

Acknowledgments. The author is indebted to Professor Steve Birch for his assistance with useful content for earlier drafts of the manuscript, along with valuable discussions. Professor Pekka Rissanen also deserves special thanks for helping provide the time and space necessary for undertaking the study. Of course, any mistakes that remain are entirely the responsibility of the author.

Financial Support. This work was supported by an unconditional grant from the Yrjö Jahnsson Foundation (grant number 6572), with the funder having no other role in the study itself, in interpretation of the results, or in the writing of the manuscript.

Conflict of Interest. The author has nothing to disclose.

References

1. **Culyer AJ** (2018) Cost, context, and decisions in health economics and health technology assessment. *Int J Technol Assess Health Care* **34**(5), 434–41.
2. **Culyer AJ, Jönsson B** (ed.) (1986) *Public and private health services: complementarities and conflicts*. Oxford: Basil Blackwell.
3. **Williams A** (1974) The cost-benefit approach. *Br Med Bull* **30**(3), 252–6.
4. **Birch S, Gafni A** (1992) Cost effectiveness/utility analyses. Do current decision rules lead us to where we want to be? *J Health Econ* **11**(3), 279–96.
5. **Coast J** (2004) Is economic evaluation in touch with society's health values? *BMJ* **329**(7476), 1233–6.
6. **Drummond M** (1980) *Principles of economic appraisal in health care*. Oxford: Oxford University Press.
7. **Drummond M, Sculpher M, Claxton K *et al.*** (2015) *Methods for the economic evaluation of health care programmes*. 4th ed. Oxford: Oxford University Press.
8. **Smith A** (1776) *An inquiry into the nature and causes of the wealth of nations*. London: Strahan and Cadell; Book I, Chapter VI, 1.
9. **Green DI** (1894) Pain-cost and opportunity-cost. *Q J Econ* **8**(2), 218–29.
10. **Coase RH** (ed.) (1977) *Economic forces at work (A collection of papers by Armen Albert Alchian)*. Indianapolis: Liberty Press.
11. **Buchanan JM** (2008) *Opportunity cost. The new Palgrave dictionary of economics*. 2nd ed. London: Palgrave Macmillan UK, 1–5.
12. **McIntosh E, Donaldson C, Ryan M** (1999) Recent advances in the methods of cost-benefit analysis in healthcare. Matching the art to the science. *Pharmacoeconomics* **15**(4), 357–67.
13. **Sloan FA, Hsieh CR** (2017) *Health economics*. 2nd ed. Cambridge, Massachusetts: MIT Press.
14. **Tianviwat S, Chongsuvivatwong V, Birch S** (2009) Optimizing the mix of basic dental services for Southern Thai schoolchildren based on resource consumption, service needs and parental preference. *Community Dent Oral Epidemiol* **37**(4), 372–80.
15. **Edlin R, McCabe C, Hulme C *et al.*** (2015) *Cost effectiveness modelling for health technology assessment: a practical course*. London: Adis.
16. **Brent RJ** (2014) *Cost-benefit analysis and health care evaluations*. 2nd edn. Cheltenham: Edward Elgar.
17. **Neumann PJ, Ganiats TG, Russell LB *et al.*** (eds.) (2016) *Cost-effectiveness in health and medicine*. 2nd ed. Oxford: Oxford University Press.
18. **Culyer AJ** (2015) Why do/should we do economic evaluation? *Value Outcomes Spotlight* **1**(2), 8–10.
19. **Karimi M, Brazier J** (2016) Health, health-related quality of life, and quality of life: What is the difference? *Pharmacoeconomics* **34**(7), 645–9.
20. **Birch S, Donaldson C** (1987) Applications of cost-benefit analysis to health care: Departures from welfare economic theory. *J Health Econ* **6**(3), 211–25.
21. **Sendi P, Gafni A, Birch S** (2002) Opportunity costs and uncertainty in the economic evaluation of health care interventions. *Health Econ* **11**(1), 23–31.
22. **Birch S, Gafni A** (2011) The inconvenient economic truth: benefits forgone as an input to economic evaluation and implications for decision-making. In: Rosen B, Israeli A, Shortell S, eds. *Improving health and healthcare who is responsible? Who is accountable?* Jerusalem, Israel: The Israel National Institute for Health Policy Research, 601–22.
23. **Nyborg K** (2014) Project evaluation with democratic decision-making: What does cost–benefit analysis really measure? *Ecol Econ* **106**, 124–31.
24. **Shackley P, Donaldson C** (2000) Willingness to pay for publicly-financed health care: how should we use the numbers? *Appl Econ* **32**(15), 2015–21.
25. **Quade ES** (1971) *A history of cost-effectiveness*. Santa Monica, CA: RAND Corporation.
26. **Franklin B** (1842) *Memoirs of Benjamin Franklin*. New York: Harper & Brothers.
27. **Simon HA** (1957) *Administrative behavior: a study of decision-making processes in administrative organization*. New York: Macmillan.
28. **Ilias G, Joanna C, Ed D *et al.*** (2016) Maximizing health or sufficient capability in economic evaluation? A methodological experiment of treatment for drug addiction. *Med Decis Making* **37**(5), 498–511.
29. **Daniels N, van der Wilt GJ** (2016) Health technology assessment, deliberative process, and ethically contested issues. *Int J Technol Assess Health Care* **32**(1–2), 10–5.
30. **Claxton K** (1999) The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* **18**(3), 341–64.
31. **Paulden M** (2016) *Opportunity cost and social values in health care resource allocation*. Alberta: University of Alberta.
32. **Hurley J** (1998) Chapter 16: Welfarism, extra-welfarism and evaluative economic analysis in the health sector. In: Barer ML, Getzen TE, Stoddart GL, eds. *Health, health care and health economics: perspectives on distribution*. Chichester: Wiley, pp. 373–95.

33. **Mooney G** (1979) Values in health care. In: Lee K, ed. *Economics and health planning*. London: Croom Helm, pp. 23–44.
34. **Culyer AJ** (2016) Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. *Health Econ, Policy Law* **11**(4), 415–32.
35. **Thokala P, Ochalek J, Leech AA, Tong T** (2018) Cost-effectiveness thresholds: the past, the present and the future. *Pharmacoeconomics* **36** (5), 509–22.
36. **Claxton K, Martin S, Soares M *et al.*** (2015) Methods for the estimation of the NICE cost effectiveness threshold. *Health Technol Assess* **19**(14), xxix–xxxiv.
37. **Birch S, Gafni A** (2006) The biggest bang for the buck or bigger bucks for the bang: the fallacy of the cost-effectiveness threshold. *J Health Serv Res Policy* **11**, 46–51.
38. **Caro JJ** (2009) Pursuing efficiency: a dead end for HTA? *Value Health* **12**, S49.
39. **Cleemput I, Neyt M, Thiry N *et al.*** (2011) Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care* **27**(1), 71–6.
40. **Birch S, Gafni A** (2016) Population needs, opportunity costs and economic methods for financial sustainability in health care systems. In: Ethgen O, Staginnus U, eds. *The future of health economics*. London: Routledge, pp. 169–180.
41. **Gafni A** (2006) Economic evaluation of health-care programmes: is CEA better than CBA? *Environ Resour Econ* **34**(3), 407–18.
42. **Mooney G** (2012) *The health of nations: towards a new political economy*. London: Zed Books.
43. **Vickers G** (1981) Systems analysis: a tool subject or judgment demystified? *Policy Sci* **14**(1), 23.
44. **Baltussen R, Jansen MPM, Bijlmakers L *et al.*** (2017) Value assessment frameworks for HTA agencies: the organization of evidence-informed deliberative processes. *Value Health* **20**(2), 256–60.
45. **Garrison Jr LP, Neumann PJ, Willke RJ *et al.*** (2018) A health economics approach to US value assessment frameworks—summary and recommendations of the ISPOR special task force report [7]. *Value Health* **21**(2), 161–5.
46. **Mauskopf JA** (1998) Prevalence-based economic evaluation. *Value Health* **1**(4), 251–9.
47. **Deaton A, Cartwright N** (2018) Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* **210**, 2–21.
48. **Manski CF** (2019) The lure of incredible certitude. *Econ Philos*, 1–30, https://doi.org/10.1017/S0266267119000105.
49. **Oortwijn W, Sampietro-Colom L, Habens F** (2017) Developments in value frameworks to inform the allocation of healthcare resources. *Int J Technol Assess Health Care* **33**, 1–7.
50. **Schultze CL** (1969) Why benefit-cost analysis? In: Hinrichs HH, Taylor GM, eds. *Program budgeting and benefit-cost analysis: cases, text and readings*. Pacific Palisades: Goodyear Publishing Company, Inc., pp. 1–8.
51. **Enthoven AC** (1966) Operations research at the national policy level. In: Tucker SA, ed. *A modern design for defense decision: a McNamara-Hitch-Enthoven anthology*. Washington: Industrial College of the Armed Forces, pp. 149–160.
52. **Marsh KD, Sculpher M, Caro JJ, Tervonen T** (2018) The use of MCDA in HTA: great potential, but more effort needed. *Value Health* **21**(4), 394–7.
53. **Culyer AJ, Bombard Y** (2012) An equity framework for health technology assessments. *Med Decis Making* **32**(3), 428–41.
54. **Williams B** (1981) Conflicts of values. In: Williams B, ed. *Moral luck: philosophical papers 1973–1980*. Cambridge: Cambridge University Press, 71–82.
55. **European network for Health Technology Assessment (EUnetHTA) project**. (2016) HTA core model: version 3.0. Available at http://www.corehta.info/model/HTACoreModel3.0.pdf. Accessed 2018.
56. **Sandel MJ** (2013) Market reasoning as moral reasoning: why economists should re-engage with political philosophy. *J Econ Perspect* **27**(4), 121–40.
57. **Mooney G** (1998) "Communitarian claims" as an ethical basis for allocating health care resources. *Soc Sci Med* **47**(9), 1171–80.
58. **Hitch CJ, McKean RN** (1960) *The economics of defense in the nuclear age*. Cambridge, Mass: Harvard University Press.
59. **Wildavsky A** (1993) *Speaking truth to power: the art and craft of policy analysis*. New Brunswick, NJ: Transaction.
60. **Mooney G** (2002) Priority setting in mental health services. *Appl Health Econ Health Policy* **1**(2), 65–74.
61. **Birch S** (1997) As a matter of fact: evidence-based decision-making unplugged. *Health Econ* **6**(6), 547–59.
62. **Stone DA** (2002) *Policy paradox: the art of political decision making*. New York: W.W. Norton.
63. **Brazier JE, Rowen D, Lloyd A, Karimi M** (2019) Future directions in valuing benefits for estimating QALYs: is time up for the EQ-5D? *Value Health* **22**(1), 62–8.
64. **Culyer AJ, Chalkidou K** (2019) Economic evaluation for health investments en route to universal health coverage: cost-benefit analysis or cost-effectiveness analysis? *Value Health* **22**(1), 99–103.
65. **Mooney G, Russell E, Weir R** (1980) *Choices for health care*. London: MacMillan, 177 p.
66. **O'Donnell R** (2016) Complexities in the examination of opportunity cost. *J Econ Educ* **47**(1), 26–31.
67. **Daddario EQ** (1967) House of Representatives Bill 6698. Washington: U.S. Govt. Print. Off., March 7, 1967.
68. **Macfie AL** (1949) What kind of experience is economizing? *Ethics* **60**(1), 19–34.
69. **Marseille E, Larson B, Kazi DS *et al.*** (2015) Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bull World Health Organ* **93**(2), 118–24.

PUBLICATION IV

Cost-effectiveness analysis of PSA-based mass screening: Evidence from a randomised controlled trial combined with register data

Neill Booth, Pekka Rissanen, Teuvo L.J. Tammela, Paula Kujala, Ulf-Håkan Stenman, Kimmo Taari, Kirsi Talala & Anssi Auvinen

PLOS ONE, 14(11), e0224479

PLOS ONE

RESEARCH ARTICLE

Cost-effectiveness analysis of PSA-based mass screening: Evidence from a randomised controlled trial combined with register data

Neill Booth[1]*, Pekka Rissanen[1], Teuvo L. J. Tammela[2,3], Paula Kujala[3,4], Ulf-Håkan Stenman[5], Kimmo Taari[6], Kirsi Talala[7], Anssi Auvinen[1]

1 Faculty of Social Sciences (Health Sciences), Tampere University, Tampere, Finland, **2** Department of Urology, Tampere University Hospital, Tampere, Finland, **3** Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland, **4** Department of Pathology, Fimlab Laboratories, Tampere, Finland, **5** Department of Clinical Chemistry and Haematology, University of Helsinki, Helsinki, Finland, **6** Department of Urology, University of Helsinki, Helsinki, Finland, **7** Finnish Cancer Registry, Helsinki, Finland

* Neill.Booth@tuni.fi

Citation: Booth N, Rissanen P, Tammela TLJ, Kujala P, Stenman U-H, Taari K, et al. (2019) Cost-effectiveness analysis of PSA-based mass screening: Evidence from a randomised controlled trial combined with register data. PLoS ONE 14(11): e0224479. https://doi.org/10.1371/journal.pone.0224479

Editor: Christopher J.D. Wallis, University of Toronto, CANADA

Received: September 2, 2019

Accepted: October 15, 2019

Published: November 5, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0224479

Data Availability Statement: Identifiable individual-level data cannot be shared publicly because of Finnish legislation governing the

Abstract

In contrast to earlier studies which have used modelling to perform cost-effectiveness analysis, this study links data from a randomised controlled trial with register data from nationwide registries to reveal new evidence on costs, effectiveness, and cost-effectiveness of organised mass prostate-cancer screening based on prostate-specific antigen (PSA) testing. Cost-effectiveness analyses were conducted with individual-level data on health-care costs from comprehensive registers and register data on real-world effectiveness from the two arms of the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC), following 80,149 men from 1996 through 2015. The study examines cost-effectiveness in terms of overall mortality and, in addition, in terms of diagnosed men's mortality from prostate cancer and mortality with but not from prostate cancer. Neither arm of the FinRSPC was clearly more cost-effective in analysis in terms of overall mortality. Organised screening in the FinRSPC could be considered cost-effective in terms of deaths from prostate cancer: averting just over one death per 1000 men screened. However, even with an estimated incremental cost-effectiveness ratio of below 20,000€ per death avoided, this result should not be considered in isolation. This is because mass screening in this trial also resulted in increases in death with, but not from, prostate cancer: with over five additional deaths per 1000 men screened. Analysis of real-world data from the FinRSPC reveals new evidence of the comparative effectiveness of PSA-based screening after 20 years of follow-up, suggesting the possibility of higher mortality, as well as higher healthcare costs, for screening-arm men who have been diagnosed with prostate cancer but who do not die from it. These findings should be corroborated or contradicted by similar analyses using data from other trials, in order to reveal if more diagnosed men have also died in the screening arms of other trials of mass screening for prostate cancer.

protection of personal data. For data access requests, interested researchers can contact the relevant Finnish registries, for example, via https://thl.fi/en/web/thlfi-en/statistics/information-for-researchers/authorisation-application.

Funding: The study was supported by the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (grant numbers 9N064 and 9R002 to NB); the Yrjö Jahnsson Foundation (grant number 6572 to NB); the Academy of Finland (grant number 260931); the Pirkanmaa Cancer Society and the Cancer Society of Finland. The funding organisations did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript and only provided financial support in the form of authors' salaries and/or research materials. Fimlab Laboratories provided support in the form of a salary for PK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: PK has taken part in a conference with support from Amgen. Fimlab Laboratories provided support in the form of a salary for PK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. TT has acted as a consultant for Orion Pharma, Bayer AG, and Ferring and received research funding from Medivation, Pfizer, and Lidds AB. KT has taken part in a conference with support from Astellas and received research funding from Medivation, Astellas, and Pfizer. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

There has been a wide range of evidence published on the effectiveness of systematic prostate-specific-antigen–based screening in reducing prostate-cancer mortality [1–4]; however, associated estimates of costs and cost-effectiveness from real-world data have received much less attention [5, 6]. Organised mass screening based on prostate-specific antigen (PSA) testing potentially offers systematic early detection of aggressive prostate cancer at a curable stage and thereby reduction of mortality [7]. However, the PSA test is not specific for cancer, as increased PSA levels can equally indicate benign changes in the prostate, so the PSA test has the potential to lead to harmful overtreatment [8]. Of course, questions extend beyond the clinical realm: policy-level ones can be asked, about what PSA-based organised screening might "cost" in relation to the "benefits" produced [9, 10]. Such relationships between costs and effectiveness (i.e., economic efficiency) are often described through some form of cost-effectiveness analysis (CEA) [11, 12]. While modelling-based CEA can provide useful information, its results typically are highly dependent on both the data and the assumptions used, which may sometimes be flawed or inaccurate [13, 14]. The need for assumptions can be minimised and the data quality maximised by drawing conclusions directly from the results of a pragmatic randomised controlled trial; we take this approach here, benefiting, e.g., from Finland's well established statutory health-care registries [15–17].

In light of the above considerations, a CEA was conducted with the primary aim of providing empirical estimates of some of the relationships between costs and effects of PSA screening from the Finnish Randomised Study of Screening for Prostate Cancer (FinRSPC) after 20 years of the trial, using intention-to-screen analysis of health-care costs, mortality, and cost-effectiveness.

Materials and methods

The FinRSPC

The complete age-based cohort for the FinRSPC was selected by staff at the Finnish population registry and consists of all men born in 1929–1944 residing in the Helsinki or Tampere region and alive on the date of randomisation (January 1 of each year from 1996 through 1999, a total of 80,458 men were randomised). Those men randomised to the screening arm were systematically invited for organised tests (serum PSA determination) at a local clinic, while those in the control arm received no such invitation as part of the trial. Three screening rounds were arranged, at four-year intervals, with men above 71 years of age no longer invited. Serum PSA was used for the primary screening test, with a cutoff of 4 ng/mL and ancillary testing for men with PSA 3.0–3.9 (digital rectal examination in 1996–1997, free/total PSA ratio from 1997 onwards). Randomisation occurred before consent, i.e., in order to prevent self-selection biases all men in the target age cohorts were randomised to one of the two arms without their consent being sought, this was undertaken in full accordance with Finnish legislation at that time. Follow-up started on January 1 in the year of randomisation and ended at death, upon emigration, or on the common closing dates for analyses of both costs and effectiveness (December 31, 2012–2015).

Register-data permissions and sources

The collection of data for this research was approved by the relevant Institutional Review Boards: by the Finnish data-protection authority, by the National Institute for Health and Welfare (THL), by Statistics Finland (TK-53-1330-18), and by the Ethics Committees of the participating university-hospital districts. The need for consent from the men assigned to

the trial was waived by a ruling from THL for the current register-based study (Official decision number: THL/36/5.05.00/2009). Data were obtained from several registries and entered in the FinRSPC database, using each man's unique Finnish personal identity code as the key for deterministic record linkage. Cancer cases were identified from the Finnish Cancer Registry (FCR), causes of death from Statistics Finland; episodes of hospital care from the THL-maintained Care Register for Health Care (CRHC), and prescription-medicine reimbursements from the nationwide register (PMRR) maintained by the Social Insurance Institution of Finland. The PMRR contains information on the exact costs of outpatient prescription medications paid by the healthcare sector in Finland. The CRHC is a comprehensive national register which covers inpatient stays in, and outpatient visits to, hospitals. To classify and identify resource use, we used the Finnish version of the Nordic Diagnosis Related Group (NordDRG) -system [18]. Identifiable individual-level data cannot be shared publicly because of Finnish legislation governing the protection of personal data. The data underlying the results presented in the study can be obtained from the relevant Finnish authorities for researchers who meet the criteria for access to confidential data. The funding organisations did not play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript and only provided financial support in the form of authors' salaries and/or research materials. Fimlab Laboratories provided support in the form of a salary for PK, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Costs

Our CEA follows a healthcare-sector perspective using register-based costs; utilising individual-level data on publicly-provided secondary and tertiary health-care visits and stays for men in the FinRSPC during the 20-year trial. In addition, the PMRR provides, to the nearest cent (¢), the costs of outpatient prescription medications paid by the Social Insurance Institution of Finland. For costs of secondary and tertiary care we used the most applicable NordDRG cost weights (in euros), which the THL had gathered from Finnish hospitals, for both inpatient and outpatient costs. The cost of the screening intervention itself was estimated by the FCR to cost approximately 50 euros per screen (including organisation of invitations, drawing of the blood sample, and PSA determinations but not any diagnostic evaluations, since the costs of diagnostic tests are captured in our other cost estimates). All results are rounded to the nearest 100 euros to yield a level of precision suitable for comparative estimates of cost and cost-effectiveness. Our base case analysis uses a discount rate of 3% per annum [19], and all euro amounts were adjusted using the most appropriate price indices available from Statistics Finland. Further details about these data sources and costs, as well as about the study design and trial registration have been provided in earlier FinRSPC or European Randomised Study of Screening for Prostate Cancer (ERSPC) publications ([3, 5, 20] or [2]).

Analyses

The register data available on both costs and effects were analysed in accordance with the intention-to-screen principle; that is, they were examined in accordance with the initial trial-arm assignment. We used mortality as the measure of effectiveness, because no other register-based effectiveness data were available (e.g., on health-related quality of life) [20]. All follow-up is truncated at 17 years, with men who were randomised on January 1 1996 were followed up until December 2012, whereas, e.g., men who were randomised on January 1 1999 were followed up until December 2015. All tests of statistical significance are two sided, with Cox proportional hazards regression used in the mortality analysis. Our CEA calculates incremental

cost-effectiveness ratios (ICERs) by means of the data from the FinRSPC and national registers on costs and effects [21]. Our health-economic approach focuses on differences in mortality, including comparisons of the numbers of men dying between the two arms [19]. These health-economic comparisons are reported in line with current standards, with the primary result reported here being the overall ICER for the FinRSPC in terms of overall mortality [13], with additional CEA analysis for two subgroups, firstly for men who died from prostate cancer and, secondly, for men who died with, but not from prostate cancer [22]. All data handling and analysis, including the merging of data from different registers, was performed using Stata [23].

Results

In all, 31,867 men were assigned to the screening arm and 48,282 to the control arm, with 3,788 men in the screening arm (11.9%) and 5,050 men in the control arm (10.5%) being diagnosed with prostate cancer, respectively (Fig 1 [24]). Data on both costs and effectiveness were recorded in the registers used in our study for 31,740 men in the screening arm (100%) and for 48,075 men in the control arm (100%). However, no cost records were found for 127 men in the screening arm and 207 men in the control arm, including one man in the screening arm who was diagnosed with and subsequently died of prostate cancer.

After 20 years of the trial, no statistically significant differences were observed between the arms in terms of the estimated average health-care costs of all men (Table 1). Although average costs for the 792 men who died of prostate cancer were around 10% higher in the screening arm (not statistically significant), there were negligible differences in total costs between the arms, since relatively few men died from prostate cancer in the screening arm. While average

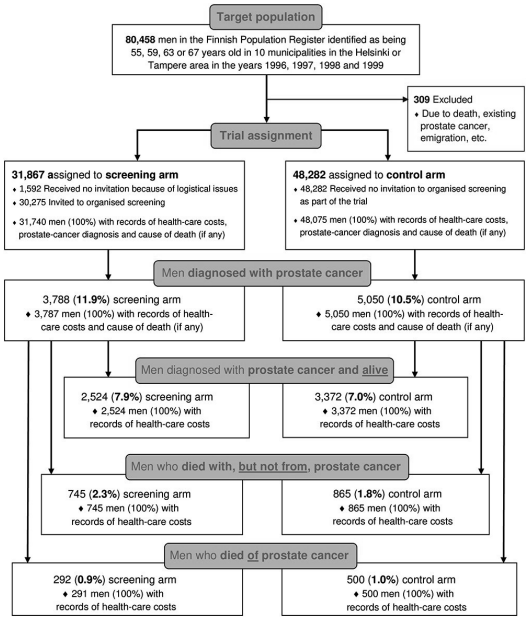

Fig 1. Enrollment and health-related outcomes.

https://doi.org/10.1371/journal.pone.0224479.g001

Table 1. Comparisons and statistical tests of the real-world health-care cost estimates. Results comparing trial arms during the 17-year follow-up.

Estimated all-cause health-care costs (register-based)	N in control arm	Mean in control arm	N in screening arm	Mean in screening arm	Difference between means (standard error)	Two-sided t-test	Difference† in total costs (in millions)
All men	48,075	**€37,800**	31,740	**€37,600**	**-€200** (€400)	p = 0.65	-€5.3
Men **not diagnosed** with prostate cancer	43,025	**€36,100**	27,953	**€35,600**	**-€500** (€400)	p = 0.26	-€29.2
Men **diagnosed** with prostate cancer	5,050	**€38,800**	3,787	**€39,300**	**€400** (€1,100)	p = 0.64	€23.9
Men who have **survived with** a prostate cancer diagnosis	3,372	**€46,300**	2,524	**€46,700**	**€300** (€1,100)	p = 0.76	€14.7
Men who **died with, but not from, prostate cancer**	865	**€62,400**	745	**€60,000**	**-€2,400** (€3,100)	p = 0.43	€9.7
Men who have **died of prostate cancer**	500	**€63,600**	291	**€68,500**	**€5,000** (€4,700)	p = 0.29	-€0.5

† = Differences are calculated as total costs in screening arm minus total costs in control arm and adjusted to take account of the relative size of the trial arms (rounded to the nearest hundred thousand euros).

https://doi.org/10.1371/journal.pone.0224479.t001

costs for the 1,610 men who died with but not from prostate cancer were approximately 5% lower in the screening arm (not statistically significant), a small substantive increase in total health-care costs for this subgroup was observed in the screening arm, as the rightmost column in Table 1 shows. This is because more men in this arm, i.e., a higher percentage of men in the screening arm, died with prostate cancer but not from it.

There was no statistically-significant difference in all-cause mortality (hazard ratio (HR) = 1.006, 95% confidence interval [CI], 0.98 to 1.03; P = 0.625) (Fig 2, Panel (A)). However, among diagnosed men there was a reduction in prostate-cancer-specific death in the screening arm: HR = 0.78, 95% CI, 0.68 to 0.90; P = 0.001 (Fig 2, Panel (B)). In addition, non-prostate-cancer mortality for men diagnosed with prostate cancer was higher in the screening arm than in the control arm: HR = 1.16, 95% CI, 1.05 to 1.27; P = 0.004 (Fig 2, Panel (C)). This increase in the rate of non-prostate-cancer mortality for men diagnosed with prostate cancer in the screening arm, seems to be most pronounced over five years after randomisation (as Fig 2, Panel (C) shows). Overall, differences in mean health-care costs and mean effectiveness for diag-nosed men between the trial arms were relatively small with regard to both prostate-cancer mortality and non-prostate-cancer mortality, with relatively high standard error (Table 2).

Our primary CEA produced a ICER which shows there was a health-related harm at less cost. This primary CEA result is presented in Fig 3, Panel (A), and shows that 95% confidence intervals are not able to be defined due to the uncertainty surrounding this estimate. Although not statistically significant, after 20 years, the impact of the FinRSPC equates to just under two additional deaths overall for every 1,000 men in the screening arm, with negligible savings in health-care costs. This finding reflects approximately 63 more deaths overall in the screening arm, which in turn reflects the negative contribution of approximately 174 more deaths observed in the screening arm for men diagnosed with prostate cancer, after having adjusted for the difference in size of the trial arms (Figs 1 and 2 and Table 2). Therefore, the ICER estimated for the FinRSPC overall, can be expressed qualitatively as a statistically non-significant reduction in costs accompanied by a statistically non-significant increase in the number of deaths (Table 2).

The estimated ICER for men diagnosed with, and who died of, prostate cancer, is 19,400€ per prostate-cancer death averted. This reflects the 38 or so fewer deaths from prostate cancer

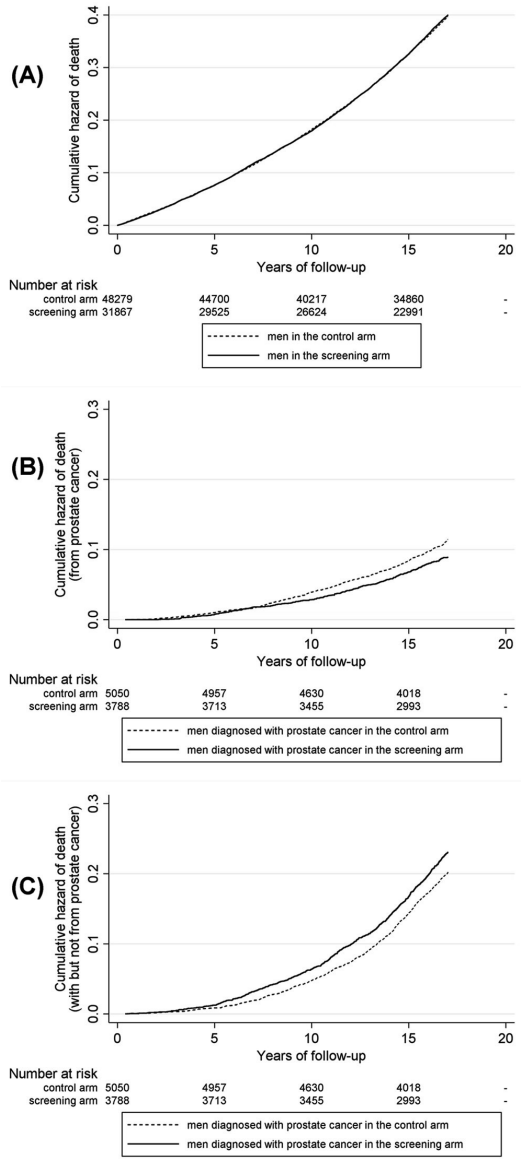

Fig 2. Nelson–Aalen estimates of risk of dying, from point of randomisation. Panel (A): Death from any cause, during follow-up, by trial arm. **Panel (B):** Death from prostate cancer, for men diagnosed during follow-up, by trial arm. **Panel (C):** Death with, but not from, prostate cancer, for men diagnosed during follow-up, by trial arm.

https://doi.org/10.1371/journal.pone.0224479.g002

being observed in the screening arm after 20 years of the FinRSPC. However, only if a decision-maker's willingness to pay per 'prostate cancer death averted' is over 120,000€ could the screening arm of the FinRSPC be considered 'cost-effective' at conventional levels of statistical significance (Fig 3, Panel (B)).

Table 2. Comparisons between the screening and control arms. Register-based health-care cost estimates, observed effectiveness and incremental cost-effectiveness ratios, after 17 years of follow-up.

	Control arm		Screening arm		Differences (screening arm—control arm)			Incremental cost-effectiveness ratio (ICER)
	Mean cost in euros	Mean effect (percentage of deaths*)	Mean cost in euros	Mean effect (percentage of deaths*)	in mean cost in euros	in mean effect (percentage of deaths†)	in number of deaths averted‡	Point estimate [effectiveness measure]
All men in the trial, using the effectiveness measure of deaths from any cause:								
Mean (**total**)	**37,800**	**0.327**	**37,600**	**0.329**	**-100**	**0.001**	**-63**	reduction in costs and increase in deaths§ [death from any cause]
(standard error)	(300)	(0.002)	(300)	(0.003)	(400)¶	(0.001)¶		
Men diagnosed with prostate cancer, using the effectiveness measure of deaths from prostate cancer:								
Mean (**total**)	**38,800**	**0.010**	**39,300**	**0.009**	**400**	**-0.02**	**38**	19,400€‖ [death from prostate cancer]
(standard error)	(600)	(<0.001)	(600)	(<0.001)	(900)¶	(0.006)		
Men diagnosed with prostate cancer, using the effectiveness measure of cause of death something other than prostate cancer:								
Mean (**total**)	**38,800**	**0.018**	**39,300**	**0.023**	**400**	**0.02**	**-174**	increase in costs and increase in deaths# [death with, but not from, prostate cancer]
(standard error)	(600)	(<0.001)	(600)	(<0.001)	(900)¶	(0.008)¶		

† = percentages expressed as decimals;

‡ = adjusted to take account of the relative size of the trial arms (rounded to the nearest integer);

§ = a 95% confidence interval is not able to be defined due to the uncertainty surrounding this estimate (see Fig 2, Panel (A));

¶ = bootstrap standard error;

‖ = an increase in mean costs (not statistically significant), and a statistically-significant increase in deaths averted (see Fig 2, Panel (B));

= an increase in mean costs (not statistically significant) and a statistically-significant reduction in deaths averted, i.e., a statistically significant increase in deaths (see Fig 2, Panel (C)).

https://doi.org/10.1371/journal.pone.0224479.t002

The analysis of cost-effectiveness above suggests a need to report on one further CEA too, this one focusing on death from other causes than prostate cancer among men diagnosed with prostate cancer (Fig 3, Panel (C)). This secondary analysis reports estimated cost-effectiveness for the men diagnosed with prostate cancer, 174 more of whom perished in the screening arm from causes other than prostate cancer, i.e., just over five additional deaths per 1000 men screened. Fig 3, Panel (C) also reflects the likelihood that these additional deaths also come at a cost in terms of health-care (of around 20,000€ per additional death).

Discussion

We examined the costs, effectiveness, and cost-effectiveness connected with a large population-based comparative-effectiveness trial of organised PSA screening for prostate cancer. Taking each of these elements in turn, firstly, costs; in terms of mean health-care costs, we found no indications of statistically-significant differences overall. However, such differences may not be discernible due to the extensive heterogeneity observed in the trial participants' utilisation of health-care services; i.e., extremely high health-care costs for some men in the trial reduced the mean estimates' ability to fully describe the cost impact [5, 25]. For example, average and total overall costs for men not diagnosed with prostate cancer were, somewhat surprisingly, somewhat lower in the screening arm, even though men in the screening arm were attributed the additional cost of screening (Table 2). This may suggest differences in health-

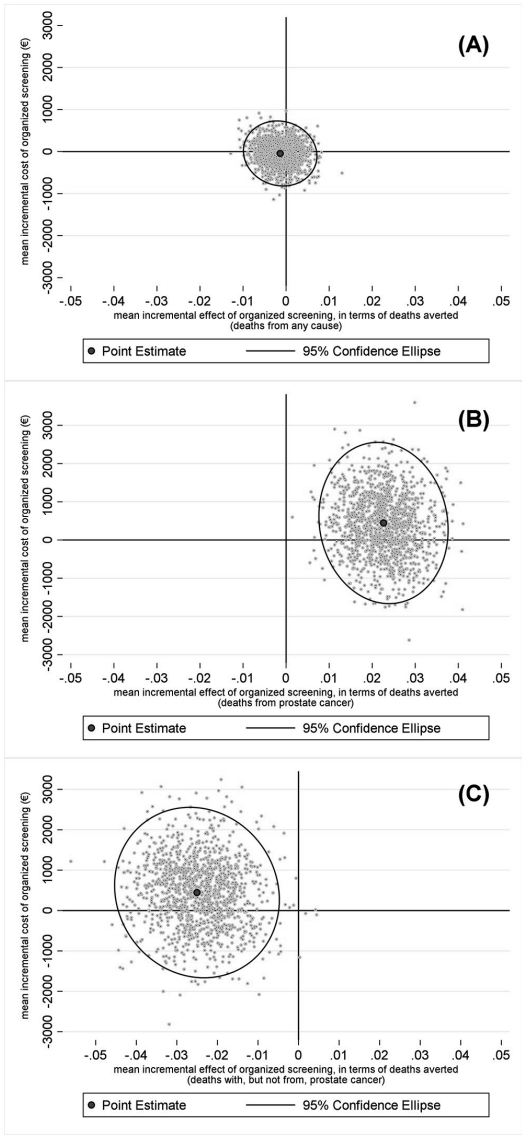

Fig 3. Scatterplots of bootstrap replications of incremental cost-effectiveness ratios in terms of the number of deaths. Panel (A): Estimates of incremental cost-effectiveness ratios in terms of death from any cause (for all men in the trial). **Panel (B):** Estimates of incremental cost-effectiveness ratios in terms of death from prostate cancer (for men diagnosed with prostate cancer). **Panel (C):** Estimates of incremental cost-effectiveness ratios in terms of death from causes other than prostate cancer (for men diagnosed with prostate cancer).

https://doi.org/10.1371/journal.pone.0224479.g003

care costs in our study may have more to do with random fluctuations or outliers in health-care costs than screening itself. Of course, ideally the time horizon necessary for a comprehensive cost-effectiveness analysis would be one that is long enough for all relevant costs (and outcomes) to manifest (see, e.g., [26] or [19]). However, our analyses did not attempt to make predictions about future costs or survival for the trial population, as robust methods were not available to extrapolate from the health-care costs and mortality effects observed for men in the FinRSPC who have already died, to those who may die in the coming years [27]. One further reason for not extrapolating data beyond the within-trial horizon, is that the follow-up period of the trial covered an expanse of time that witnessed many changes in prostate-cancer treatment protocols.

Secondly, effectiveness, in exploratory mortality analyses prompted by our CEA findings in terms of all-cause mortality and prostate-cancer mortality, we also evaluated comparative effectiveness in terms of non-prostate-cancer mortality among the men diagnosed with prostate cancer. We undertook these secondary analyses because it was apparent from our other analysis results that, although on average prostate-cancer mortality was lower in the screening arm among men diagnosed with the disease, all-cause mortality was higher in the screening arm overall (even though this latter result was not statistically significant). To determine whether or not our CEA findings in terms of all-cause mortality were due to chance and given that the main impact of screening is, *a priori*, likely to be upon men diagnosed with prostate cancer, we undertook further analyses of mortality among men diagnosed with prostate cancer. The above secondary analyses are not, of course, undertaken in full accordance with the intention-to-screen (ITS) principle, however, three results reported in Fig 1, which do use the ITS principle, should be noted. The first result of note is that, relative to the control arm, on average 13% more men were diagnosed with prostate cancer in the screening arm (11.9% versus 10.5%). The second result of note is that, relative to the control arm, on average 13% less men died from prostate cancer in the screening arm (0.9% versus 1.0%). The third result of note in Fig 1 and, perhaps, the most important one here is that, relative to the control arm, on average 27% more men in the screening arm died with, but not from, prostate cancer (2.3% versus 1.8% in the control arm). A partial explanation for the result that, relative to the control arm, on average 27% more men in the screening arm died with, but not from, prostate cancer, could be related to overdiagnosis; with 13% more men on average diagnosed in the screening arm, relative to the control arm. Such a 'labelling' effect could plausibly account for approximately half of the observed additional non-prostate cancer deaths in diagnosed men. A second plausible explanation could be linked to competing causes of death; if more men are on average spared from death from prostate cancer due to PSA mass screening they may die of other causes. However, analysis using a proxy for survival time–i.e., the follow-up time in each arm–as the outcome measure (instead of number of deaths) also revealed an overall decrease in 'survival time' in the screening arm, for men diagnosed with prostate cancer. Put together, however, even the possible explanations listed above would still only seem to partially explain the finding of higher mortality for screening-arm men who have been diagnosed with prostate cancer but who do not die from it.

Although increased cardiovascular mortality due to endocrine therapy or the fear or stigmatization associated with cancer diagnosis may play some role in our findings about non-prostate-cancer mortality in men diagnosed with prostate cancer, such explanations remain only speculations as to why more men in the screening arm died with prostate cancer but not from it. In addition, we were not able to identify any single specific cause of death (or groups of causes of death), such as deaths related to intentional self-harm, or other underlying differences between the arms, which could explain these mortality differences. The most marked increases in non-prostate-cancer mortality were among those men in the screening arm with

Tumor-Node-Metastasis -stage T1c cancers at diagnosis (i.e., impalpable cancers detectable only by PSA testing [28]). Although our data suggest mid-level socioeconomic status may have been associated with an increase in non-prostate-cancer mortality when men were diagnosed with prostate cancer at any other than stage than stage T1c, these findings were not statistically significant. Possibly due to the relatively small number of observed deaths at this stage, such adjustments for socioeconomic status had minimal material impact on the differences between the arms in terms of mortality among men with prostate cancer. Further explanatory analysis is beyond the scope of this exploratory study.

Thirdly, we turn to the results relating to the cost-effectiveness of organised screening in the FinRSPC, which varied according to the outcome measure used. We present three ICERs: there was negligible impact of mass screening in the FinRSPC in terms of death from any cause in all men, (what can be interpreted as) a positive impact for death from prostate cancer in diagnosed men, and (what can be interpreted as) a negative impact for death with, but not from, prostate cancer in diagnosed men. Sensitivity analysis showed that using a discount rate of 5% and 1% does not result in major changes in the differences in costs or cost-effectiveness between the two arms in any of these analyses. In their assessment of the cost-effectiveness of screening, epidemiological studies have focused mainly on disease-specific mortality [2, 29, 30], often to the exclusion of any other effects on mortality [31]. In contrast to earlier CEAs [32–36] our approach to health-economic evaluation considers not merely prostate-cancer mortality; but characterizes all-cause mortality too, along with non-prostate-cancer mortality in men diagnosed with prostate cancer. One potential pitfall in modelling cost-effectiveness in a manner which does not adequately question the underlying epidemiology is that, accordingly, any errors in the choice of outcome measures may be compounded in the act of modelling. If the choice of outcome measure is restricted by epidemiological convention this may obscure relevant effects of the intervention, resulting in models neglecting to include a potentially relevant health state, such as non-prostate-cancer mortality in men diagnosed with prostate cancer. Incorporating all potentially important mortality impacts should be seen as central in any health-economic evaluation [37].

The interpretation of incremental cost-effectiveness ratios is not a straightforward matter, as the process of interpretation is typically specific to both the ICER's content and the decision-making context in question. In the field of health-economic evaluation, a cost-effectiveness ratio usually represents some indicator of the amount of health *gained* divided by some estimate of the financial costs associated with that estimated 'health' *gain*. Only relatively infrequently, as in the secondary analysis presented here, does the cost-effectiveness ratio represent some indicator of the amount of health *lost* divided by some estimate of the financial costs of that estimated 'health' *lost*. When we report that, for men diagnosed with prostate cancer, the estimated health-care cost per additional death is around 20,000€, this means that the data suggests the trial was economically efficient at increasing non-prostate-cancer mortality for those men. It should be clear from the results presented here that cost-effectiveness ratios can contain or omit a wide range of factors. For this reason, understanding the content of each incremental cost-effectiveness ratio is important when they are interpreted, for example, how well costs and health effects are measured and analysed, and what costs and effects are, or are not, included in the analysis. In Table 3 we set out the main research assumptions and key components which underpin the health-economic evaluation of the FinRSPC.

In practice, interpretation of cost-effectiveness information requires understanding of the components and qualities of that information [39]. Table 3 is intended to provide a useful starting point for interpretation of the information about costs, effects and cost-effectiveness provided by our study [40]. Interpretation of CEA results is also usually influenced by the interplay between the decision-making context and the specific information provided by the

Table 3. Main assumptions. Key elements of the health-economic evaluation of the FinRSPC.

A) Key elements of the analysis related to the FinRSPC:	
strengths	• Assignment to the trial arms occurred without prior consent (but with the permission of the authorities) • The men randomised represented the whole target population of the Tampere and Helsinki areas during the period, i.e., all registered male citizens of the selected age groups were included • Long-term register-based follow-up was available for practically all men (>99.9%)

limitations	• Neither Finnish registries or the trial database includes consistent follow-up of either many of the possible health-related impacts, or some of the costs, associated with prostate-cancer screening
	• The FinRSPC is limited by its context, e.g.:
	  • clinical practice today may be quite different to that of the late 1990s
	  • PSA testing became more prevalent in the population over the period of the trial, which seems likely to have had a significant effect on the impacts of the screening intervention [38]
	• The long duration of follow-up may also mean that more influences unrelated to the screening trial are reflected in its results, i.e., that there is more 'noise' in the data
	• Clinical trials such as the FinRSPC typically can only provide robust information on average treatment effects for the whole trial population. This is also the case for this trial, which practically precludes robust analysis by, e.g., geographical- or age-related–subgroup

B) Key elements of the analysis related to costs:	
strengths	• The analysis uses well-established registers covering both use of hospital services (inpatient and outpatient) as well as reimbursements for almost all outpatient prescription medications • The registers provide almost complete coverage of these (hospital and prescription-medication) costs for almost all men in the trial for almost the whole duration of the 20-year study
limitations	• In principle, ideally all costs associated with PSA mass screening for prostate cancer and its consequences might be included as part of a cost-effectiveness analysis, at least when attempting to gauge the robustness of the results to the inclusion or omission of a range of cost items. • Although information relating to primary care costs is typically included, such information was not readily available from Finnish registers or the trial database, so is not included here • Various cost drivers, such as costs to patients, costs which fall on the social-care budget, and costs of lost productivity in the economy, were not included in our analyses • Although the registers provide an identical source of data for men in both arms of the trial, and although price indices and discount rates were applied uniformly in both arms, the register-based cost estimates presented here are based on NordDRG cost weights, the cost estimates are, at best, merely rough indicators of the magnitude of the true current costs which might be associated with PSA mass screening for prostate cancer
C) Key elements of the analysis related to health-related outcomes:	
strengths	• The analysis presented here focuses on one of the most important and robust impacts related to health outcomes, i.e., mortality • The analysis uses data from well-established registries and precise cause-of-death registers with practically complete coverage (at least for men who did not emigrate)
limitations	• No direct measurement of health-related quality of life or patient satisfaction was possible using the available register data
D) Key elements related to the cost-effectiveness analysis *per se*:	
strengths	• Each incremental cost-effectiveness ratio (ICER) presented looks at a different aspect of mortality, together the three ICERs presented provide an a variety of useful indicators of the efficiency of mass screening for prostate cancer in terms of the main impacts on mortality • Although the cost drivers used in our analysis are limited in scope, as noted above (in section **B)** of the table), the data provides almost complete coverage of two main costs: hospital care and prescription medications
limitations	• Each incremental cost-effectiveness ratio (ICER) presented looks at a different aspect of mortality, none of the ICERs alone provide an all-encompassing indicator of the efficiency of organised PSA mass screening for prostate cancer • As noted above (in section **C)** of this table) our analysis does not incorporate health-related quality of life considerations or considerations relating to patient satisfaction. For this reason the ICER estimates presented here provide only a truncated representation of the efficiency of PSA mass screening for prostate cancer and do not take into account important effects, e.g., on quality of life

https://doi.org/10.1371/journal.pone.0224479.t003

incremental cost-effectiveness ratios in question. Therefore, judgment will typically be needed, in every separate case and context, to gauge to what extent any estimated incremental cost-effectiveness ratios provide an indication of 'value'. Many elements of the chosen approach to health-economic evaluation can markedly influence the results of cost-effectiveness analyses [41]. When CEAs are based on a single randomised controlled trial, CEAs naturally are heavily dependent on that source of information. Although randomised controlled trials are typically seen as one of the best research methods to inform public health policy, it should be noted that they do have their weaknesses [25, 42, 43]. Further details about the strengths and limitations of our study will be set out below.

Strengths of the study

Our register-based cost-effectiveness analysis combines the power of a randomised controlled trial with extensive follow-up via real-world data from comprehensive health-care registers. By avoiding reliance on many of the assumptions typically necessary for modelling costs and outcomes, our study represents a potentially significant application of CEA to improve the knowledge base about organised screening for prostate cancer. Although numerous modelling-based studies have been reported upon [32–36, 44], their estimates or forecasts typically do not proceed from data alone, with a frequent cascading effect wherein cost estimates are based on previous estimates of outcomes. Our results can be regarded as a groundbreaking contrast, in that this is the first report on CEA based on real-world data derived from one study of PSA mass screening. The men in the FinRSPC, i.e., in the Finnish arm of the ERSPC, were a complete age cohort of the men in and around two main Finnish conurbations, Helsinki and Tampere. The men were assigned to the two arms before randomisation, thus minimising problems associated with selection to either group. Although the FinRSPC does not provide a perfectly valid assessment of organised screening versus no screening, it likely provides a potentially valid assessment of organised screening versus current clinical practice. The contamination by opportunistic PSA-testing experienced during this trial is more likely to be generalizable to current clinical practice than would 'no screening', by providing evidence of the likely impact of organised screening over and above opportunistic testing [19]. In addition, the data over the 20 years of the trial (with 17-year median follow-up time), from fairly comprehensive data on health-care costs and on effectiveness (in terms of mortality), help in obtaining potentially generalizable cost-effectiveness estimates, which realistically account for the diluting effects of contamination.

Truncation of the follow-up at 17 years was undertaken to limit our analysis to only the most robust data, because as age cohorts were selected from the population the on January 1 of each year from 1996 through 1999, analysis without truncation would mean that follow-up beyond 17 years would only be possible for fewer and fewer men each year until the maximum of 20 years of follow-up. Truncation in this study produces more conservative results, with analysis using all available data producing both larger effect sizes and more statistically-significant associations.

Another strength of our study lies in its ability to inform current and future choices of suitable metrics for effectiveness [45]. For instance, our finding that over the 20 years of the FinRSPC trial, mortality from causes other than prostate cancer among diagnosed men increased in the screening arm has potentially significant implications for future research. Our analysis likely provides a useful building block in that its findings could be input for testing existing cost-effectiveness models' sensitivity to new information, which has been shown to be useful elsewhere [46]. In addition, similar analyses from comparable trials of organised screening could provide illuminating corroboration or contradiction of the findings presented here,

because we are not aware of any other published analyses of mortality from causes other than prostate cancer among diagnosed men in trials of PSA mass screening for prostate cancer [47].

Limitations of the study

As set out in parts of Table 3, this study is limited in scope for a number of reasons, e.g., that no direct analysis of health-related quality of life or patient satisfaction was possible using the available register data [48]. The estimates from a pragmatic trial in Finland presented here are not necessarily indicative or representative of the impacts of organised PSA screening likely in other health-care systems. Indeed, these estimates are unlikely to represent exact health-care costs, effectiveness (in terms of mortality), or cost-effectiveness in other settings. However, this is unavoidable for any pragmatic long-term, real-world study. The generalizability of our analysis to other settings is dependent on how well the manner of implementing the screening intervention and subsequent care pathways in the FinRSPC can be generalized and also on contextual elements such as the treatment patterns and the relative homogeneity of the FinRSPC participants (the vast majority being Finnish and Caucasian). More generally, the health-care system in which the trial took place (in the largely publicly-funded Finnish health-care system) may limit generalizability. Despite these limitations, this trial of comparative effectiveness does represent an important source of evidence, which can be used to supplement earlier evidence from modelling studies. Of course, the authors acknowledge that, just as modelling-based CEA depends on assumptions that may sometimes be flawed or inaccurate, the relevance of empirical CEA to a wider population or time horizon also depends on assumptions that may sometimes be flawed or inaccurate. In addition, all our results should be interpreted in consideration of the likelihood of high levels of contamination in the control arm [49], since most of the men in the control arm had a PSA test at some point in the trial and the cumulative incidence of T1c cancers was, for example, only approximately 20% higher in the screening arm than in the control arm. Although almost 75% of men in the screening arm of the FinRSPC participated in the organised screening, we cannot know exactly which men underwent non-systematic screening, i.e., opportunistic testing [38]. Further, the results presented here for the subgroups of men diagnosed with prostate cancer are not necessarily causally linked to randomisation, they are the result of randomisation followed by diagnosis, so the intention-to-screen-analytic comparison between the arms is uncertain in this respect [25]. However, the results of the intention-to-screen analysis show mortality effects of similar magnitude to those in the subgroup analysis presented here. It should also be noted here that health-economic evaluations are information-intensive in their input requirements and that their use often suffers from a lack of appropriate information [50], especially as pathways to health are quite complex [51]. For example, the main costs analysed were from secondary and tertiary care, so various cost drivers, such as costs of lost productivity to the economy, costs of primary-care treatment, and costs due to social care, were not considered. One of these, primary-care costs, were not included in our study as data were not available from registers for the majority of the follow-up period (except to the extent that prescription medication use, as part of primary care, was covered). However, we are not aware of any strong reason why primary care costs would differ substantially between the groups when, e.g., there were no major differences in the costs of secondary or tertiary care. One further limitation (or, conversely, potential strength), is that our study employed a fixed time horizon, setting it apart from many model-based studies, which vary the time horizon modelled. On the other hand, models that attempt to estimate "lifetime" costs and effectiveness typically rely on assumptions that could seem out of place in light of the findings presented here, especially since we cannot know with any certainty which direction the impacts of PSA screening will take next.

Conclusion

Our primary analyses showed no major difference in overall health-care costs or in overall mortality within the 17 years of follow-up. However, in further analysis, relatively minor reductions in prostate-cancer mortality at the expense of increased costs in the screening arm were found, but these may be outweighed by an increase in mortality from other causes for men diagnosed with prostate cancer in the screening arm. Our analysis could be usefully supported or contradicted by similar analyses using data from comparable trials of mass screening. Longer-term follow-up may also allow more robust conclusions as to the balance of the benefits and harms of introducing organised PSA mass screening.

Acknowledgments

The authors would like to express thanks to all those who have helped in collating the trial data over the past 20 years but especially to Liisa Määttänen, formerly of the Finnish Mass Screening Registry, for her continued support related to the trial database, even in retirement. The corresponding author would also like to thank Jani Raitanen and Pasi Aronen for numerous helpful discussions about matters statistical and analytical. Any mistakes that may remain are the responsibility of the corresponding author.

Author Contributions

Conceptualization: Neill Booth, Pekka Rissanen, Teuvo L. J. Tammela, Anssi Auvinen.

Data curation: Neill Booth, Ulf-Håkan Stenman, Kirsi Talala.

Formal analysis: Neill Booth, Anssi Auvinen.

Funding acquisition: Neill Booth, Pekka Rissanen, Teuvo L. J. Tammela, Kimmo Taari, Anssi Auvinen.

Investigation: Neill Booth, Paula Kujala, Kimmo Taari, Kirsi Talala.

Methodology: Neill Booth, Pekka Rissanen, Paula Kujala, Anssi Auvinen.

Project administration: Neill Booth, Kirsi Talala.

Supervision: Pekka Rissanen, Anssi Auvinen.

Validation: Neill Booth, Anssi Auvinen.

Visualization: Neill Booth.

Writing – original draft: Neill Booth, Anssi Auvinen.

Writing – review & editing: Neill Booth, Pekka Rissanen, Teuvo L. J. Tammela, Paula Kujala, Ulf-Håkan Stenman, Kimmo Taari, Kirsi Talala, Anssi Auvinen.

References

1. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Ciatto S, Nelen V, et al. Prostate-Cancer Mortality at 11 Years of Follow-up. New England Journal of Medicine. 2012; 366(11):981–90. https://doi.org/10.1056/NEJMoa1113135 PMID: 22417251
2. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet. 2014; 384(9959):2027–35. https://doi.org/10.1016/S0140-6736(14)60525-0 PMID: 25108889
3. Kilpeläinen TP, Tammela TL, Malila N, Hakama M, Santti H, Määttänen L, et al. Prostate cancer mortality in the Finnish Randomized Screening Trial. Journal of the National Cancer Institute. 2013; 105(10):719–25. https://doi.org/10.1093/jnci/djt038 PMID: 23479454

4. Pinsky PF, Miller E, Prorok P, Grubb R, Crawford ED, Andriole G. Extended follow-up for prostate cancer incidence and mortality among participants in the Prostate, Lung, Colorectal and Ovarian randomized cancer screening trial. BJU Int. 2019; 123(5):854–60. https://doi.org/10.1111/bju.14580 PMID: 30288918
5. Booth N, Rissanen P, Tammela T, Taari K, Talala K, Auvinen A. Costs of screening for prostate cancer: Evidence from the Finnish Randomised Study of Screening for Prostate Cancer after 20-year follow-up using register data. European Journal of Cancer. 2018; 93:108–18. https://doi.org/10.1016/j.ejca.2018.01.111 PMID: 29501976
6. Krahn M, Zagorski B, Laporte A, Alibhai SMH, Bremner KE, Tomlinson G, et al. Healthcare costs associated with prostate cancer: estimates from a population-based study. BJU International. 2010; 105 (3):338–46. https://doi.org/10.1111/j.1464-410X.2009.08758.x PMID: 19594734
7. Yates DR, Anderson JB. Screening for prostate cancer. In: Tewari A, editor. Prostate Cancer: A Comprehensive Perspective. London: Springer-Verlag; 2013. p. 333–46.
8. Haines IE, Ablin RJ, Miklos GLG. Screening for prostate cancer: time to put all the data on the table. BMJ. 2016; 353. https://doi.org/10.1136/bmj.i2574 PMID: 27226459
9. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. New England Journal of Medicine. 1977; 296(13):716–21. https://doi.org/10.1056/NEJM197703312961304 PMID: 402576
10. Russell LB, Sinha A. Strengthening Cost-Effectiveness Analysis for Public Health Policy. American Journal of Preventive Medicine. 2016; 50(5, Supplement 1):S6–S12. https://doi.org/10.1016/j.amepre.2015.11.007 PMID: 27102861
11. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied Methods of Cost-effectiveness Analysis in Health Care. Gray A, Briggs A, editors. Oxford: Oxford University Press; 2011.
12. Neumann PJ, Kim DD, Trikalinos TA, Sculpher MJ, Salomon JA, Prosser LA, et al. Future Directions for Cost-effectiveness Analyses in Health and Medicine. Med Decis Making. 2018; 38(7):767–77. https://doi.org/10.1177/0272989X18798833 PMID: 30248277
13. Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. Gray A, Briggs A, editors. Oxford: Oxford University Press; 2014.
14. Noordzij MA, Blanker MH. Re: Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. Journal of the National Cancer Institute. 2015; 107(6):djv110. https://doi.org/10.1093/jnci/djv110 PMID: 25888716
15. Auvinen A. Prostate cancer screening: What can we learn from randomised trials? Translational Andrology and Urology. 2018; 7(1):12–7. https://doi.org/10.21037/tau.2017.12.13 PMID: 29594015
16. Ramsberg J, Neovius M. Register or electronic health records enriched randomized pragmatic trials: The future of clinical effectiveness and cost-effectiveness trials? Nordic Journal of Health Economics. 2017; 5(1):62–76. https://doi.org/10.5617/njhe.1386
17. Sackett DL. The arrogance of preventive medicine. CMAJ. 2002; 167(4):363–4. PMID: 12197692
18. The Nordic Classification Centre. NordDRG Users' Manual 2012 [19/06/19]. http://www.nordcase.org/eng/materials/manuals/.
19. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II–an ISPOR Good Research Practices Task Force report. Value in Health. 2015; 18(2):161–72. https://doi.org/10.1016/j.jval.2015.02.001 PMID: 25773551
20. Booth N, Rissanen P, Tammela TLJ, Määttänen L, Taari K, Auvinen A. Health-related quality of life in the Finnish Trial of Screening for Prostate Cancer. European Urology. 2014; 65(1):39–47. https://doi.org/10.1016/j.eururo.2012.11.041 PMID: 23265387
21. European Network for Health Technology Assessment (EUnetHTA). Methods for health economic evaluations—a guideline based on current practices in Europe. http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/2015-04-29-ECO-GL_Final%20version_0.pdf. 2015.
22. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)–explanation and elaboration: A Report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. Value in Health. 2013; 16(2):231–50. https://doi.org/10.1016/j.jval.2013.02.002 PMID: 23538175
23. StataCorp LLC. Stata statistical software: Release 15.1. College Station, TX. 2019.
24. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. Ann Intern Med. 2010; 152(11):726–32. https://doi.org/10.7326/0003-4819-152-11-201006010-00232 PMID: 20335313
25. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Social Science & Medicine. 2018; 210:2–21. https://doi.org/10.1016/j.socscimed.2017.12.005 PMID: 29331519

26. Basu A, Maciejewski ML. Choosing a Time Horizon in Cost and Cost-effectiveness Analyses. JAMA. 2019; 321(11):1096–7. https://doi.org/10.1001/jama.2019.1153 PMID: 30789668

27. Goodwin P, Wright G. The limits of forecasting methods in anticipating rare events. Technological Forecasting and Social Change. 2010; 77(3):355–68. https://doi.org/10.1016/j.techfore.2009.10.008

28. Sobin LH, Wittekind C, Gospodarowicz MK. TNM Classification of Malignant Tumours. Oxford: Wiley-Blackwell; 2010.

29. Tsodikov A, Gulati R, Heijnsdijk EAM, Pinsky PF, Moss SM, Qiu S, et al. Reconciling the Effects of Screening on Prostate Cancer Mortality in the ERSPC and PLCO Trials. Ann Intern Med. 2017; 167(7):449–55. https://doi.org/10.7326/M16-2586 PMID: 28869989

30. Editorial. DIfferent effects of screening on prostate cancer death in two trials. Annals of Internal Medicine. 2017.

31. Prasad V, Lenzer J, Newman DH. Why cancer screening has never been shown to 'save lives'–and what we can do about it. BMJ. 2016; 352. https://doi.org/10.1136/bmj.h6080 PMID: 26740343

32. Roth JA, Gulati R, Gore JL, Cooperberg MR, Etzioni R. Economic analysis of prostate-specific antigen screening and selective treatment strategies. JAMA Oncology. 2016; 2(7):890–8. https://doi.org/10.1001/jamaoncol.2015.6275 PMID: 27010943

33. Shteynshlyuger A, Andriole GL. Cost-Effectiveness of Prostate Specific Antigen Screening in the United States: Extrapolating From the European Study of Screening for Prostate Cancer. The Journal of Urology. 2011; 185(3):828–32. https://doi.org/10.1016/j.juro.2010.10.079 PMID: 21239021

34. Pataky R, Gulati R, Etzioni R, Black P, Chi KN, Coldman AJ, et al. Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada. International Journal of Cancer. 2014; 135(4):939–47. https://doi.org/10.1002/ijc.28732 PMID: 24443367

35. Shin S, Kim YH, Hwang JS, Lee YJ, Lee SM, Ahn J. Economic evaluation of prostate cancer screening test as a national cancer screening program in South Korea. Asian Pac J Cancer Prev. 2014; 15(8):3383–9. https://doi.org/10.7314/apjcp.2014.15.8.3383 PMID: 24870726

36. Heijnsdijk EAM, de Carvalho TM, Auvinen A, Zappa M, Nelen V, Kwiatkowski M, et al. Cost-effectiveness of prostate cancer screening: A simulation study based on ERSPC data. Journal of the National Cancer Institute. 2015; 107(1). https://doi.org/10.1093/jnci/dju366 PMID: 25505238

37. Mullahy J. Health and evidence in health economics. Health Economics. 2019; 28(10):1163–5. https://doi.org/10.1002/hec.3926 PMID: 31264292

38. Kilpeläinen TP, Pogodin-Hannolainen D, Kemppainen K, Talala K, Raitanen J, Taari K, et al. Estimate of opportunistic prostate specific antigen testing in the Finnish Randomized Study of Screening for Prostate Cancer. The Journal of Urology. 2017; 198(1):50–7. https://doi.org/10.1016/j.juro.2017.01.048 PMID: 28104375

39. Booth N. On value frameworks and opportunity costs in health technology assessment. International Journal of Technology Assessment in Health Care. 2019. https://doi.org/10.1017/S0266462319000643 PMID: 31530332

40. Sullivan SD, Mauskopf JA, Augustovski F, Jaime Caro J, Lee KM, Minchin M, et al. Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. Value in Health. 2014; 17(1):5–14. https://doi.org/10.1016/j.jval.2013.08.2291 PMID: 24438712

41. Bertram MY, Lauer JA, De Joncheere K, Edejer T, Hutubessy R, Kieny M-P, et al. Cost-effectiveness thresholds: pros and cons. Bulletin of the World Health Organization. 2016; 94(12):925–30. https://doi.org/10.2471/BLT.15.164418 PMID: 27994285

42. Krauss A. Why all randomised controlled trials produce biased results. Annals of Medicine. 2018; 50(4):312–22. https://doi.org/10.1080/07853890.2018.1453233 PMID: 29616838

43. Manski CF. The lure of incredible certitude. Economics and Philosophy. 2019:1–30. https://doi.org/10.1017/S0266267119000105

44. Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer: A decision analytic view. JAMA. 1994; 272(10):773–80. PMID: 7521400

45. Quade ES. Introduction and overview. In: Goldman TA, editor. Cost-effectiveness analysis: new approaches in decision-making. New York: Praeger; 1967. p. 1–16.

46. Laxy M, Wilson ECF, Boothby CE, Griffin SJ. Incremental Costs and Cost Effectiveness of Intensive Treatment in Individuals with Type 2 Diabetes Detected by Screening in the ADDITION-UK Trial: An Update with Empirical Trial-Based Cost Data. Value in Health. 2017; 20(10):1288–98. https://doi.org/10.1016/j.jval.2017.05.018 PMID: 29241888

47. Heijnsdijk EAM, Csanádi M, Gini A, ten Haaf K, Bendes R, Anttila A, et al. All-cause mortality versus cancer-specific mortality as outcome in cancer screening trials: A review and modeling study. Cancer Medicine. 2019; 8(13):6127–38. https://doi.org/10.1002/cam4.2476 PMID: 31422585

48. McNaughton-Collins M, Walker-Corkery E, Barry MJ. Health-related quality of life, satisfaction, and economic outcome measures in studies of prostate cancer screening and treatment, 1990–2000. J Natl Cancer Inst Monogr. 2004;(33):78–101. https://doi.org/10.1093/jncimonographs/lgh016 PMID: 15504921

49. Shoag JE, Mittal S, Hu JC. Reevaluating PSA Testing Rates in the PLCO Trial. New England Journal of Medicine. 2016; 374(18):1795–6. https://doi.org/10.1056/NEJMc1515131 PMID: 27144870

50. Mooney G. Priority setting in mental health services. Applied Health Economics and Health Policy. 2002; 1(2):65–74. PMID: 14619253

51. Birch S. As a matter of fact: evidence-based decision-making unplugged. Health Economics. 1997; 6(6):547–59. PMID: 9466138