Author:
Keeney, Edna

Title:
Methods for the cost-effectiveness modelling of screening interventions in an uncertain landscape

Application to screening for prostate cancer

Author:
Keeney, Edna

Title:
Methods for the cost-effectiveness modelling of screening interventions in an uncertain landscape: Application to screening for prostate cancer

Methods for the Cost-effectiveness Modelling of Screening Interventions in an Uncertain Landscape: Application to Screening for Prostate Cancer

---

By

Edna Keeney

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Health Sciences

Bristol Medical School

July 2023

ABSTRACT

Decisions need to be made about who and how to screen for diseases to optimise health in the population. Cost-effectiveness analyses of screening interventions can be associated with many areas of uncertainty due to a constantly changing landscape in screening methods, diagnostic tests, treatments and understanding of natural history. A failure to account for such uncertainty may result in incorrect or poorly informed decisions. Prostate cancer is an example of a disease where recent developments in the understanding of who and how to screen have provided challenges to the analyst trying to make recommendations on the most cost-effective screening strategy.

Using prostate cancer screening as a case study, this dissertation explores methods to handle uncertainty when modelling the cost-effectiveness of screening interventions in an uncertain landscape. The dissertation shows how a systematic review of previous models can identify areas of parameter and structural uncertainty, how to gain expert consensus with respect to relevant screening strategies, and how to appropriately adapt and calibrate an existing natural history model to a new setting.

It will demonstrate how the 22 studies identified in the systematic review informed the structure and data parameters of the natural history model and how a modified-Delphi process identified prostate cancer screening strategies that were deemed relevant by experts, including risk-stratified and adaptive approaches. It will also show how a decision model was adapted and calibrated to UK data to find that, of the strategies identified in the Delphi, a once-off screening at age 50 years was most cost-effective.

Many methods are available for dealing with uncertainty in cost-effectiveness modelling of screening interventions. The dissertation will conclude with a discussion on the merits and limitations of the methods used, with recommendations given for practice. The aim is to provide a guide to identifying and addressing the uncertainty inherent in cost-effectiveness analyses of screening strategies.

DEDICATION AND ACKNOWLEDGEMENTS

I would like to dedicate this dissertation to the following people:

First and foremost, to Rupert and Leo. Rupert for being a constant source of motivation and support, and Leo for arriving halfway through the process, almost derailing it altogether, but ultimately giving me a reason to carry on.

To my parents, for always supporting and encouraging me at every step of my education, no matter the end result.

A special mention goes to my supervisors – Richard, Howard, Sabina, and Emma. Their consistent advice and guidance over the five-year period have been excellent and invaluable in shaping this dissertation.

I would like to express my deep appreciation to Mark Clements for his generosity in sharing his time and expertise when adapting the Prostata model. I would also like to thank Roman Gulati for sharing his knowledge on the model.

To all the individuals who participated in the Delphi process, I am grateful for your expert advice and valuable contributions, which greatly influenced the final outcome of this research.

Lastly, I extend my heartfelt thanks to the University of Bristol, which provided me with the opportunity to take this significant step and where I experienced many fulfilling years of learning and growth.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .....  ..... DATE: 17/7/23

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PUBLICATIONS ARISING FROM THIS WORK

Below is a list of publications that have arisen from research undertaken throughout this PhD along with the contributions section from each publication.

The following publication is based on Chapter 3

Keeney E, Thom H, Turner E, Martin RM, Morley J, Sanghera S. Systematic Review of Cost-Effectiveness Models in Prostate Cancer: Exploring New Developments in Testing and Diagnosis. Value Health. 2022 Jan;25(1):133-146. doi: 10.1016/j.jval.2021.07.002.

Author contributions: Keeney, Thom, Turner, Martin, and Sanghera were involved in the concept and design of the study. Keeney and Morley were involved in data collection. All authors were involved in data analysis and interpretation. Keeney wrote the paper with all authors contributing to critical revisions. All authors read and approved the final manuscript.

The following publication is based on Chapter 4

Keeney E, Thom H, Turner E, Martin RM, Sanghera S. Using a Modified Delphi Approach to Gain Consensus on Relevant Comparators in a Cost-Effectiveness Model: Application to Prostate Cancer Screening. Pharmacoeconomics. 2021 May;39(5):589-600. doi: 10.1007/s40273-021-01009-6.

Author contributions: All authors were involved in the conception of the study. Keeney designed and administered the questionnaire and analysed the results with guidance from Martin, Turner, Thom and Sanghera. Sanghera, Thom, Turner and Martin contributed to the interpretation of the data and writing of the manuscript. Keeney wrote the first draft of the

manuscript with Sanghera, Thom, Turner and Martin critically revising for important intellectual content. All authors read and approved the final manuscript.

The following publication is based on Chapter 5

Keeney E, Sanghera S, Martin RM, Gulati R, Wiklund F, Walsh EI, Donovan JL, Hamdy F, Neal DE, Lane JA, Turner EL, Thom H, Clements MS. Cost-Effectiveness Analysis of Prostate Cancer Screening in the UK: A Decision Model Analysis Based on the CAP Trial. Pharmacoeconomics. 2022 Dec;40(12):1207-1220. doi: 10.1007/s40273-022-01191-1.

Author contributions: Keeney, Sanghera, Martin, Turner, Thom and Clements contributed to the study conception and design. Clements developed the model and assisted with the analyses. Keeney performed the analyses with guidance from Sanghera, Martin, Gulati, Wiklund, Turner, Thom and Clements. Walsh assisted with data provision. The first draft of the manuscript was written by Keeney and Clements and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

CHAPTER 1. INTRODUCTION

1.1. Uncertainty in cost-effectiveness analysis of screening interventions

Screening tests aim to identify asymptomatic patients harbouring specific conditions. They can be offered to the population as a whole or only to patients known to be at risk. The goal of screening is to identify people who are likely to benefit from further testing or treatments to reduce the risk of the disease or its complications.[1] National screening programmes for certain types of cancer, which aim to identify early stage disease before symptoms develop, such as breast, cervical and colorectal cancer screening, are particularly common. A seminal paper by Wilson and Jungner[2] highlighted principles that should be considered when making a decision about whether to provide a screening programme:

Wilson and Jungner's principles of screening:

- The condition sought should be an important health problem.
- The natural history of the condition, including development from latent to declared disease, should be adequately understood.
- There should be a recognizable latent or early symptomatic stage.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- There should be an agreed policy on whom to treat as patients.
- There should be an accepted treatment for patients with recognized disease.
- Facilities for diagnosis and treatment should be available.
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case-finding should be a continuing process and not a "once and for all" project.

The most reliable way to evaluate many of these principles, and determine the effectiveness of screening interventions, is to randomise asymptomatic individuals into screening and no screening programmes in a randomised controlled trial and evaluate outcomes and costs after a sufficiently long follow up period.[1 3] The National Screening Committee in the UK state

that before a screening programme is introduced, "There should be evidence from high quality randomised controlled trials that the screening programme is effective in reducing mortality or morbidity".[4] However, as often occurs, the screening outcomes e.g. increase in life expectancy, only become clear after a long period of time. Within that time period the medical landscape can change with new and rapid medical advances common, particularly in the case of cancer diagnosis and treatment.[5 63] Once screening trials have begun it is often impossible or impractical to alter the screening or treatment protocol as implementing trial amendments can be associated with a substantial amount of time and resources, particularly when large numbers of patients have been randomised.[7] Despite this, decisions need to be made about whether screening programmes should be made available by national healthcare payers such as the United Kingdom (UK) National Health Service (NHS), and in what format. Advances and innovations in screening and diagnostic tools need to be considered when making these decisions.

Providing a national screening programme requires considerable healthcare resources. Economics is the study of how society allocates its scarce resources among alternative uses. The goal of health economics is to ensure efficiency in a health system by maximising benefits from the healthcare resources available.[8] As there are always limited resources, due to a limited healthcare budget, choices must be made about which healthcare interventions are made available by the healthcare payer. Health economists have developed the tools of economic evaluation to establish whether the current set of services provided in the health sector is appropriate or whether there is an alternative set which would improve health in the population.[9] Economic evaluations commonly try to assess the cost-effectiveness of a new technology, which is whether the benefits incurred, compared to the existing medical strategy, are worth the additional expenses.

Often a single study or trial will not provide all evidence needed to carry out an economic evaluation. In addition, as the timeframe of an analysis needs to be long enough to capture all relevant differences in costs and outcomes between an intervention and its comparators, it is common that trials do not have sufficient follow up to estimate this directly. Another issue is that relevant alternatives may not have been directly compared in trials. Economic decision analytic modelling is a method used to represent possible consequences from

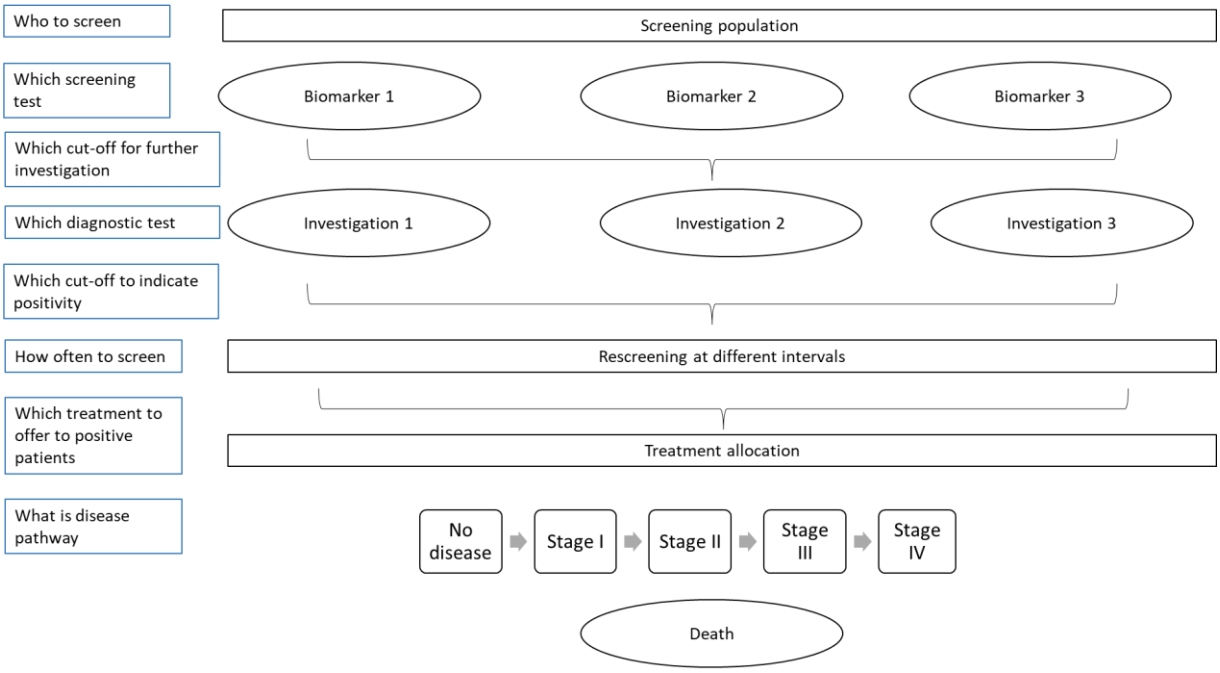
implementing an intervention, when such uncertainty exists. Modelling allows relevant data on the effectiveness of interventions, costs and outcomes, which may come from different sources, to be synthesized using appropriate statistical methods.[10] Model-based economic evaluation is therefore a way of predicting the long-term effects and costs of novel screening strategies, when long-term trial results are not available.[11]

The first step in any economic evaluation is to determine the decision question, which includes defining the population of interest, the intervention being evaluated and its relevant comparators.[12] The cost-effectiveness of a screening strategy may be heavily dependent on the population chosen in terms of risk factors such as age and ethnicity.[13 14] Cost-effectiveness will also depend on the frequency of screening, screening test, diagnostic test or tests used, treatment allocation, and their order and combination, which are all subject to changing practice.[1] This can make identifying relevant screening strategies or comparators in an economic evaluation of screening programmes challenging.

Another potential issue in the case of screening is a lack of understanding surrounding the natural history of the condition in question, for example the rate of progression of a tumour, and how this is impacted by screening. This is particularly problematic as those who are not screened are often not followed up. In practice, the pathway of these people before they were detected is unknown, as is what would have happened if they were not detected and treated. The impact of screening on disease progression with the addition of new screening strategies or treatments is also uncertain. This is a common problem in model-based economic evaluation of screening interventions as many parameters on the natural progression of potentially malignant lesions are not directly observable.[15]

Figure 1 demonstrates the different areas of uncertainty in the economic evaluation of screening interventions, including the initial population to screen, the tests and diagnostic strategies to use, the rescreening process, which treatments to offer and how to allocate these by disease severity, and alongside this, uncertainty in the natural history of the disease and the impact screening will have on progression and mortality. This highlights the challenges that arise for an analyst trying to make recommendations on the cost-effectiveness of screening strategies.

Figure 1. Areas of uncertainty in economic evaluation of screening interventions



1.2. Case study: Prostate cancer screening

Prostate cancer is an example of an important health problem with a recognisable latent stage (where men have the disease but are asymptomatic) that could be a candidate for a national screening programme. Although it is the leading cause of male cancer in the UK, no national screening programme is currently in place.[16] It is also an area where the medical landscape has undergone many recent changes in terms of testing, diagnosis and treatment.[17-20] The effectiveness and cost-effectiveness of prostate cancer screening is currently unclear with trial evidence varying[21-23] and modelling studies showing varying estimates of overall quality of life impact and cost-effectiveness.[24] Standard diagnostic methods lead to overdetection of cancers that may not progress to become clinically important in a man's lifetime, but can also miss aggressive, potentially fatal prostate cancer.[25 26] This has resulted in poor consensus and conflicting evidence on the screening strategies that are currently relevant, or whether screening should be undertaken at all. Recent research suggests, however, that tailored screening according to a man's predicted risks may improve the cost-effectiveness of screening.[27 28 19]

Further uncertainty exists as to how the natural history of prostate cancer should be modelled. In prostate cancer, models have generally characteristed prostate cancer progression by Gleason grade (≤6, 7, or ≥ 8), cancer stage (T1-2, T3-4, M1), or both, which are measures of how quickly the cancer is growing and if it has spread outside the prostate. A paper by Sanghera et al[24] reviewed the literature and identified nine cost-effectiveness models in PSA-based prostate cancer screening with each capturing the natural history of the condition and the impact of screening on disease progression to differing degrees of detail. Differences ranged from one model considering only presence or absence of cancer[29] to another differentiating between 18 pre-clinical and clinical stages.[30] Sanghera et al suggested that this may be due to disagreement or changing trends overtime on how prostate cancer progression should be represented.

1.3. Objectives of PhD

The aim of this PhD is to provide a guide to identifying and dealing with the uncertainty that is inherent in cost-effectiveness modelling of screening strategies, using prostate cancer screening as a case study.

Specific objectives are to:

1. Carry out a systematic literature review to assess the evidence base on recent cost-effectiveness models which have considered new innovations in prostate cancer screening (Chapter 3).
2. Explore methods to gain consensus on relevant screening and diagnostic strategies by getting agreement from clinicians, modellers, experts in prostate cancer and other relevant stakeholders (Chapter 4).
3. Adapt and calibrate a cost-effectiveness model for use in a UK setting (Chapter 5).
4. Use agreed strategies and calibrated model to carry out a model-based economic evaluation of prostate cancer screening strategies (Chapter 6).
5. Reflect on methods used and make recommendations for future analysts (Chapter 7).

CHAPTER 2. BACKGROUND

2.1. Screening programmes

2.1.1. What is a screening programme

The seminal publication by Wilson & Jungner[2] stated:

"Screening is the presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment."

As such, screening programmes involve not only the screening test but also the subsequent diagnosis, treatment and follow up of the disease. The World Health Organization outline the key stages in a screening programme[31], which include:

1. Identifying the population eligible for screening
2. Inviting them to be screened, including providing information to help inform their decision on whether to accept the invitation
3. Conducting the screening test
4. Referring people with positive results for further testing and informing participants with screen negative results
5. Diagnosing true cases and identifying false positives
6. Treating, following up and monitoring cases
7. Analysing and reporting outcomes to improve the screening programme

Adjustments can be made at each of these stages that may improve the effectiveness and cost-effectiveness of a screening programme.

2.1.2. Test accuracy measures

No screening test is entirely precise, meaning that some individuals who do not have the condition may get a positive test outcome (known as a false positive), while some individuals with the condition may get a normal or negative outcome (known as a false negative). Test accuracy measures indicate the ability of the test to distinguish between people who do have the condition (true positives) from those that do not (true negatives). The measures used to indicate how well a test performs are termed sensitivity and specificity. The sensitivity of a test is its ability to correctly identify people *with* the condition and specificity is its ability to correctly identify people *without* the condition.[32] The formula to calculate sensitivity is therefore the number of true positives identified out of the total number of true positives in the population, and the formula for specificity is the total number of true negatives identified out of the total number of true negatives in the population. The threshold or value chosen as the cut-off between a positive and negative result can affect whether a screening test is considered more sensitive and less specific or vice versa.

Other measures frequently used in screening are the positive predictive value (PPV) and negative predictive value (NPV). The PPV is the proportion of patients with positive test results who are correctly diagnosed, while the NPV is the proportion of patients with negative test results who are correctly diagnosed.[33] The estimation of sensitivity, specificity, PPV and NPV are vital to establishing the true effect of a screening programme as the outcomes will depend on the ability of a screening test to accurately identify people with the disease.

2.1.3. Harms of screening

If a screening programme works well, it can prevent ill health or death and improve overall quality of life in society. Screening can also have unintended harms, however, such as

testing-related adverse events and overdiagnosis. Testing-related adverse events such as anxiety, or complications associated with diagnostic interventions such as biopsy, are particularly problematic when a screening test results in a high number of false positive results. This can result in healthy people being subjected to unnecessary tests and their associated complications, as well as increasing costs for the health system. A high number of false negative results is also undesirable as this can cause patients with the condition to ignore symptoms as they believe themselves to be healthy, thus delaying diagnosis.[34] For this reason, screening tests must aim to be as accurate as possible in terms of both sensitivity and specificity.

Another harm associated with screening is overdiagnosis. This relates to the detection of abnormalities that would never have caused harm or resulted in symptoms for the individual's remaining life.[35] The issue is that once a condition such as cancer is identified, treatment or monitoring with tests and biopsies must be offered, even though it may not be needed. The risk of overdiagnosis is higher in older age groups, as comorbidities increase and life expectancy decreases. Vickers et al suggested that stopping prostate cancer screening at age 70 could reduce overdiagnosis by 42%, for example.[36] All potential screening programmes must aim to identify those individuals for whom the potential benefits outweigh these potential harms.

2.2. Prostate cancer screening

2.2.1. Prostate cancer

The prostate is a part of the male reproductive system located between the bladder and the rectum. Its purpose is to produce fluid that makes up a part of semen.[37] The prostate also produces a protein called prostate specific antigen (PSA) that is responsible for making the semen easier to expel.[38] Prostate cancer develops when abnormal cells start to grow in the prostate gland. Prostate cancer is one of the most frequently diagnosed cancers worldwide and in many countries is the leading cause of cancer death.[39] In the UK, it is the second most common cause of cancer death in males (14% of all cancer deaths, 2017-2019).[16] Most men are diagnosed because they present with symptoms such as increased urinary

frequency, having to strain when urinating, and an inability to completely empty the bladder, caused by swelling of the prostate gland. [40]

2.2.2. Prostate cancer staging and survival

If a man develops prostate cancer, then a Gleason score/grade and cancer stage (T1-2, T3-4, N1, M1) are assigned at cancer onset. In cancer stages T1-2, cancer is usually slow growing, the tumour is found only in the prostate and PSA levels are medium or low. In stages T3-T4, PSA levels can be high, the tumour is growing and the cancer is likely to have spread beyond the prostate. N1 indicates that the tumour has spread to nearby lymph nodes and M1 indicates that it has spread to distant parts of the body.[41]

The Gleason grade of a cancer refers to how much the cancer cells look like normal cells. The overall Gleason score is estimated taking into consideration the two most common Gleason grades found in the cancer cells. One score is assigned to the most predominant pattern of cells in a biopsy and another to the second most predominant pattern. The grades are then added together to determine the Gleason score e.g., 3 + 4 = 7. Table 1 shows the Gleason grade group definitions as described by Cancer Research UK.[42] A Gleason grade of 6 or less (grade group 1) is considered a low-grade cancer, 7 (grade group 2-3) is medium-grade, and 8 or more (grade group 4-5) is high-grade. A lower-grade cancer grows more slowly and is less likely to spread than a high-grade cancer.[43] [44]

Survival is generally high for men diagnosed with prostate cancer, with almost 80% of men in the UK surviving for 10 years or more.[45] A recent UK study (ProtecT) found that after 15-years of follow-up, death from prostate cancer occurred in only 2.7% of 1643 men diagnosed with localized prostate cancer.[20] Once the cancer has spread, however, survival is lower, with only 50% of men with metastatic prostate cancer surviving for 5 years or more after diagnosis.[45]

Table 1. Cancer Research UK grade group definitions

Gleason score	Grade Group	What it means
Gleason score 6 (or 3 + 3 = 6)	Grade Group 1	The cells look similar to normal prostate cells. The cancer is likely to grow very slowly, if at all
Gleason score 7 (or 3 + 4 = 7)	Grade Group 2	Most cells still look similar to normal prostate cells. The cancer is likely to grow slowly
Gleason score 7 (or 4 + 3 = 7)	Grade Group 3	The cells look less like normal prostate cells. The cancer is likely to grow at a moderate rate
Gleason score 8 (or 4 + 4 = 8)	Grade Group 4	Some cells look abnormal. The cancer might grow quickly or at a moderate rate
Gleason score 9 or 10 (or 4 + 5 = 9, 5 + 4 = 9 or 5 + 5 = 10)	Grade Group 5	The cells look very abnormal. The cancer is likely to grow quickly

2.2.3. Treatment

To avoid the harms of treatment, active surveillance is recommended for men with low-risk (localized) prostate cancer that is found early and growing slowly. This involves monitoring the cancer and only beginning treatment if it shows signs of progressing. Other treatment options for cancer that has not spread outside the prostate include surgery or radiation therapy. For higher-risk cancers, treatment options include hormonal therapy, also called androgen deprivation therapy (ADT), alongside radiation therapy. Metastatic prostate cancer is often treated with ADT, occasionally in combination with other therapies. If a cancer stops responding to ADT, it is termed castration-resistant prostate cancer. Such advanced cancers may be treated with chemotherapy, immunotherapy, radiation therapy, or other novel interventions.[46]

2.2.4. Risk factors

Known risk factors for developing prostate cancer include older age, ethnicity and family history. Incidence rates tend to increase from age 45 onwards with the highest rates found in the 75-79 age group.[47] In terms of ethnicity, a study using prostate cancer incidence and

mortality data for England (2008–2010) by major ethnic group found that Black men are twice as likely to be diagnosed with, and die from, prostate cancer than White men. In contrast, Asian men are significantly less likely to be diagnosed compared to White men.[48] Family history and genetics have also been shown to be influential with one meta-analysis estimating the pooled rate ratio of developing prostate cancer for a man with a father or brother with prostate cancer to be 2.48 (95% confidence interval: 2.25–2.74).[49] Genome-wide association studies have identified more than 200 loci associated with prostate cancer development which account for an estimated 34–43% of the relative risk relating to family history.[50-53]

2.2.5. Goal of screening

The aim of prostate cancer screening is to identify high-risk localised prostate cancer and treat it before it has spread beyond the prostate, therefore preventing the ill health and death associated with advanced disease.[39]

2.2.6. The Prostate-specific antigen test and Transrectal Ultrasound-Guided biopsy

The most common screening methods used in prostate cancer have been the Prostate-specific antigen (PSA) test and Transrectal Ultrasound-Guided (TRUS) biopsy. The PSA test, which has been in common use as a screening tool since the 1990s, is a simple and cheap blood test which measures the level of PSA in the blood. A higher-than-average PSA can indicate if men need further investigation for prostate cancer. However, it is known to have poor accuracy, with a high incidence of false positive and false negative results.[54 55] A finding of a high PSA level has typically led to a TRUS biopsy to confirm diagnosis. However, the process is associated with infection and other adverse effects and approximately 70–80% of biopsies following a positive PSA test are negative.[56] False negative results are found in up to 25% of cases. [57 58] Therefore, although these methods may be used to catch potentially high-risk cancers that have yet to develop, there are disadvantages, including a chance of false diagnosis and overdiagnosis. Such diagnoses can lead to a series of tests and treatments that may cause unnecessary harm such as urinary incontinence and erectile dysfunction.[59]

2.2.7. Current prostate cancer screening recommendations

Recommendations vary across the world and are often updated to reflect new evidence. In recent years there has been a general consensus that screening should only be offered based on shared decision making between a man and his physician.[60] However, in 2022 the European Union recommended that organised screening programmes should be considered for prostate cancer.[61]

In summary:

- There is currently no prostate cancer screening programme in place in the UK; however, men over 50 can request a PSA test after receiving information about the advantages and disadvantages of testing.[62]
- The American Cancer Society recommends that men have a discussion with their healthcare provider about the potential benefits and risks of prostate cancer screening at age 50 for those at average risk, and earlier for those at higher risk, including African American men and those with a family history of the disease. If they decide to be screened, they are offered a PSA test. If their PSA is less than 2.5 ng/mL it is recommended that they be retested every 2 years and if it is above this level yearly testing is recommended.[63]
- The US Preventive Services Task Force also recommends that men have a discussion with their healthcare provider about the potential benefits and risks of prostate cancer screening before deciding to have a PSA test. They recommend screening from 55 to 70 years with no screening in men above this age bracket.[39]
- The American Urological Association recommends against PSA screening in men under age 40 years, over 70 years (unless they are in excellent health), in average risk men under 55, and in any man with a life expectancy less than 10-15 years. Shared decision making is recommended for men aged 55 to 69 and the use of tools such as urinary and serum biomarkers, imaging, and risk calculators should be considered prior to biopsy in men with a suspicious PSA level. A screening interval of two years or more, individualized by baseline PSA level, is recommended.[64]

- The European Association of Urology[59] recommends that men aged 50 years (younger if they have a family history of prostate cancer, are of African descent, or are carrying BRAC2 mutations), with a life expectancy of greater than 10 years, should be offered PSA testing every 2-4 years (for those with a PSA value of 1-3 ng/ml) or 5 years (for those with a PSA value of < 1 ng/ml and < 60 years old). Those over 60 years old with a PSA < 1 ng/ml should have no further testing.
- The Council of the European Union currently recommends that countries should evaluate the feasibility and effectiveness of the implementation of organised screening programmes including PSA testing in combination with magnetic resonance imaging (MRI).[61]

2.2.8. Previous trials

Table 2 summarises the key prostate cancer screening trials to date and highlights the variation found in the primary outcome of a reduction in death caused by prostate cancer (prostate cancer-specific mortality). The European Randomised study of Screening for Prostate Cancer (ERSPC), conducted across eight European countries, found that repeated PSA screening significantly reduced prostate cancer-specific mortality by 20% at 16 years of follow-up (rate ratio 0.80, 95% confidence interval 0.72–0.89). There was variation in screening frequency across countries with intervals varying from 2 years in Sweden to 7 years in Belgium, with most countries using a 4-year interval.[65] There was also variation in the PSA cut-off used for further investigation, with some centres using 3 ng/ml and some using 4 ng/ml.[66 67] A screening trial of annual PSA testing conducted in Quebec between 1988 and 1999 found that prostate cancer-specific mortality was 62% lower in the screened group after 11 years of follow-up. [68]

However, other trials have failed to show a significant mortality benefit associated with screening. The UK Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) trial[69] of a single PSA screen found no significant difference in prostate cancer mortality after a median follow-up of 10 years. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening trial, where men received annual PSA tests, failed to show a reduction in prostate cancer-specific mortality at 15 years of follow-up. However, the results from PLCO have

been questioned as most control-arm men (86%) in this trial underwent some PSA testing and almost half received annual screening.[70]

Table 2. Summary of prostate cancer screening trials

Trial	Enrolment period	Setting	N randomised to screening	Ages of men screened	Screening frequency	Mortality difference in screened and non-screened arms
Cluster Randomized Trial of PSA Testing for Prostate cancer (CAP)[69]	2001 – 2009, 10 years follow up	UK	195,912	50-69	Single screen	None
European Randomized Study of Screening for Prostate cancer (ERSPC)[21]	1993 – 2003, 16 years follow up	Europe	72,891	55-69	Every 2-7 years	21% reduction
Prostate, Lung, Colorectal, and Ovarian Cancer Screening (PLCO)[22]	1993 – 2001, 15 years follow up	US	38,340	55-74	Annually	None
Lundgren et al 2018[71]	1988 – 2003, 20 years follow-up	Sweden	2,400	55-70	Single screen	None
Labrie et al 2004[68]	1988 – 1999, 11 years follow-up	Canada	31,133	45-80	Annually	61.5% reduction

2.2.9. Overdiagnosis in prostate cancer

The results from previous trials highlight the potential for prostate cancer screening to increase cancer detection without a proportional reduction in advanced-stage disease or mortality. Autopsies conducted on elderly men in the USA, who died from causes unrelated to prostate cancer, revealed that 36% of white men and 51% of black men aged 70-79 had prostate cancer.[72 73] This indicates the significant presence of slow-growing or non-aggressive tumours in asymptomatic individuals. Therefore, the development of more sensitive tests that can identify additional prostate cancers, without distinguishing between indolent and aggressive cancer, may result in more harm than good by subjecting a larger number of patients to unnecessary interventions.[74]

2.2.10. Potential screening strategies

Since the initiation of these trials over twenty years ago, there have been significant changes in diagnostic strategies and understanding of which men should be screened. It is now becoming recognised that tailored screening according to a man's predicted risks may improve the effectiveness of screening.[27 28] Techniques to achieve this include:

- Biomarker tests to replace or complement PSA-based testing e.g. PHI, 4Kscore, SelectMDx and PCA3 (described in Appendix 1). These have been developed which act as additional reflex tests to aid the decision about when a man should be referred for biopsy.
- Using polygenic risk scores, which indicate genetic susceptibility to disease, to determine the population to be screened and intensity of screening.[27 75-77]
- Alternatives to standard TRUS biopsy, including the use of multiparametric magnetic resonance imaging (mpMRI) as a pre-biopsy triage test, to guide biopsy, or as a replacement for PSA. This approach might allow men with no or likely indolent cancer to avoid an unnecessary biopsy and improve diagnostic accuracy for more aggressive disease.[25 78 79] Alternatives to TRUS biopsy are described in Appendix 1.

These advancements offer opportunities for improving the outcomes of prostate cancer screening, particularly reduction in overdiagnosis and higher specificity for potentially lethal prostate cancer. However, the optimal combination of screening population, test, diagnostic strategy, and interval to next screen is unclear.

2.3. Economic evaluation

2.3.1. Role of economic evaluation in health care

Drummond defined economic evaluation as "the comparative analysis of alternative courses of action in terms of both costs (resource use) and consequences (outcomes, effects)".[80] Economic evaluation is needed in healthcare to determine whether healthcare resources are being used efficiently, i.e. achieving the best value for money.[81] Efficiency in healthcare relates to both technical efficiency and allocative efficiency. Technical efficiency considers how best to achieve a particular objective such as which type of hip replacement to offer to those who need it, whereas allocative efficiency considers where and how much resources should be allocated across the health system. Allocative efficiency, for example, would consider whether hip replacements should be offered at all or whether resources would be better spent elsewhere, such as on physiotherapy services.[82]

2.3.2. Quality-adjusted Life Years

A common way of measuring the effects of a healthcare intervention in economic evaluation is through the use of Quality-adjusted Life Years (QALYs). QALYs take into account both the length of time a patient spends in particular states of health and the quality of life experienced during that time, by assigning a numerical value, which is anchored between 0 (death) and 1 (perfect health) to each health state. Negative QALYs are also possible whereby a health state is considered to be worse than dead.[83] In this way QALYs consider not only whether a new intervention prolongs life compared to its alternative but also whether the quality of that life is improved. The use of QALYs enables comparison across all interventions and areas of healthcare.[84]

QALYs are calculated by multiplying the time spent in a health state by the quality of life or utility score associated with that health state. Utility scores can be estimated using direct methods or indirect methods. Direct methods ask people to value a particular health state e.g., advanced prostate cancer, using methods such as the time trade-off approach. This approach involves presenting individuals with two different scenarios and asking them to choose between them. In one scenario, they would live for the remainder of their life in the state of impaired health (advanced prostate cancer) and in the other they would live for a shorter duration but in full health. The duration of time spent in full health is adjusted until the individual reaches a point where they are indifferent between the two choices. Another type of direct method is the standard gamble, which presents individuals with a choice between staying in a specific health state or taking a gamble that could result in either full health or the possibility of death. The probability of experiencing death in the gamble is adjusted until the individual reaches a point of indifference between the certainty of the current health state and the uncertain outcome of the gamble. [84]

Indirect methods involve the use of generic preference-based measures which describe health states using standardized questionnaires, covering general aspects of health. An example of this is the EuroQol (EQ)-5D which has five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Respondents are asked to assign a level of severity ranging from “no problems” to “severe problems” to each dimension.[85] Each possible combinations of answers can then be converted into a utility score. Different countries use different value sets to convert questionnaire responses into utility scores based on values derived from the general population. Other commonly used generic questionnaires include the Short Form 6D (SF-6D)[86] and the Health Utilities Index (HUI).[87] Condition-specific instruments, which may be more sensitive in detecting effects in certain disease areas, are also commonly used.[88]

The National Institute for Health and Care Excellence (NICE) provides guidance to the NHS in England on the cost-effectiveness of interventions. As different methods used to measure health-related quality of life produce different utility values, NICE have specified that the EQ-5D is their preferred method of measuring utilities and should be used unless a case can

be made that it is inappropriate. This ensures consistency across analyses and facilitates decision making.[89]

2.3.3. Types of Economic Evaluation

There are four types of economic evaluation: cost-utility, cost-effectiveness, cost-benefit, and cost-consequence analyses. Cost-utility analyses measure costs in monetary terms and benefits as QALYs. Cost-effectiveness analyses also measure costs in monetary terms but express benefit in natural units such as hospital admissions avoided or reduction in pain. The terms cost-utility analysis and cost-effectiveness analysis are often used interchangeably to mean a cost-utility analysis.

An alternative approach is cost-benefit analysis, where both benefits and costs are measured in monetary terms.[90] In this type of analysis individuals' preferences are elicited via what they are willing to pay (or give up) for the outcomes of the healthcare intervention in question.[91 92] The advantage of cost-benefit analysis is that, once benefits have been converted into monetary terms, it can be used to compare the net economic benefit of activities both within and outside the healthcare sector, such as education and transport. However, due to methodological difficulties in carrying out this type of analysis[93], cost-utility analysis remains the most widely used, and recommended by organisations such NICE.

A final approach is cost-consequence analysis, where outcomes and costs are presented in a disaggregated manner. All manner of effects can be presented, including both health and non-health. The aim is to give decision makers a comprehensive summary of the different costs and effects, and allow them to draw their own conclusions.[94]

2.3.4. Perspectives

Economic evaluations are most often carried out from the perspective of the healthcare payer. The costs considered are therefore costs directly associated with the healthcare system including hospitalizations, medical visits including GP and outpatient attendances, medical procedures and pharmaceuticals. A societal perspective might also be taken whereby the costs of an intervention are considered from not only a healthcare system's

perspective but also from the broader societal perspective, including, for example, private costs incurred by patients and caregivers. In this case all direct and indirect costs associated with the intervention will be taken into account, including productivity losses and informal care costs.[95]

2.3.5. Presenting results of economic evaluations

Once the costs and benefits of a new intervention and its comparator/s are established, cost-effectiveness analyses relate the differences in costs between the options being compared to the differences in benefits (e.g. QALYs). If an intervention costs less than its alternatives and results in greater benefits then it is cost-effective. In these situations health economists would say that the intervention dominates its alternative.[96] However, as is often the case, if an intervention costs more than its comparators but also generates greater benefit, then the cost per benefit gained must be estimated. This can be expressed as the incremental cost-effectiveness ratio (ICER) or the ratio of incremental costs to effectiveness outcomes.

To determine whether a new intervention is cost-effective or provides good value for money, ICERs are generally compared against a cost-effectiveness threshold. Interventions with an ICER below a threshold are considered cost effective, while those with an ICER above the threshold are not.[97] The threshold is therefore an indication of the maximum amount a healthcare payer is willing to pay for one QALY gained. NICE recommend funding interventions with an ICER below a threshold of £20,000 or £30,000 per QALY gained[98], however this ranges widely across countries and there is disagreement on the empirical basis for such thresholds.[99] [100]

The results of cost-effectiveness analyses are often presented on the cost-effectiveness plane, with effects on the horizontal axis and costs on the vertical axis. The efficiency frontier is constructed by plotting the mean cost and effect of each intervention being compared and forming a line connecting each non-dominated option. Interventions above this line are not considered to be cost-effective. Cost-effectiveness acceptability curves are another means of presenting results which reflect the probability of an intervention being

most cost-effective at each willingness to pay per QALYs gained threshold. The probability of each intervention being optimal at each threshold is estimated by counting the proportion of samples for which the expected net benefit is highest. The expected net benefit is calculated by multiplying QALYs by the willingness to pay threshold and subtracting costs (mean QALYs × willingness to pay threshold – mean costs).[101]

Barton et al demonstrate that, although cost-effectiveness acceptability curves can represent decision uncertainty, they can also be misleading as the strategy with the highest probability of being cost-effective is not always the strategy with the highest expected net benefit. This is a particular problem when multiple interventions are compared. It is therefore recommended that the cost-effectiveness acceptability frontier, which shows the probability that the strategy with the highest expected net benefit is cost-effective as a function of WTP[102], always be presented alongside cost effectiveness acceptability curves, as well as an estimate of the expected value of information (described in section 2.6.4).[103]

2.3.6. Welfarism vs. Extra-welfarism

Different frameworks for carrying out economic evaluations exist. The primary frameworks in use are welfarism and extra-welfarism. Although definitions are not clear cut, welfarism typically relates to the idea that the output of healthcare should be judged according to the extent to which it maximises the individuals' perceived value of the welfare that results from it, whereas the extra-welfarist framework concentrates on maximising health.[104] Using QALYs or outcomes specific to the disease area as the measure of benefit in a cost-effectiveness analysis is considered an extra-welfarist approach whereas cost-benefit analysis is generally considered a welfarist approach.

2.4. Decision modelling as a vehicle for economic evaluation

Economic decision analytic modelling is a method used to represent possible consequences from implementing an intervention when data on the effectiveness of interventions, costs

and outcomes over the relevant timeframe is not available from a single trial and needs to be synthesised.[10]

2.4.1. Types of models

There are various types of models and the one chosen should depend on the decision question. Cohort models consider the population as a whole, with proportions under-going different events, while individual level models consider individuals with specific attributes, sampling each one individually. Both follow progress in patients over time.[105]

Decision trees are a type of cohort model that are often used for modelling relatively simple and short-term decision questions. Decision trees consist of decision nodes where decisions are made on the allocation of interventions, chance nodes where patients have a probability of experiencing particular events as a result of this allocation, and terminal nodes where the costs and outcome measures of interest, such as QALYs, are attached to the pathway taken by the patient. The proportion of patients taking each pathway is used to calculate overall costs and effects.[106]

Markov models are a more sophisticated type of cohort model where events are modelled as transitions from one health state to another. These transitions occur at the end of each model cycle which is a period of time, e.g. one month or one year, chosen to represent clinical meaningfulness. Patients move between states until they enter an absorbing state, such as the 'dead' state. As in decision trees, movement between states is dependent on transition probabilities.[10] Each state is associated with costs and outcomes and the amount of time spent by patients in each health state determines overall results. Markov models are more appropriate for situations in which events occur over a long period of time and often repeat. A limitation of Markov models is that transition probabilities are not influenced by the pathway taken to a particular health state or length of time spent in a health state.[106]

Discrete event simulation or microsimulation models are individual level models which allow event rates to be influenced by previously experienced events by simulating the impact of interventions on individuals, rather than estimating a mean response for a homogeneous

cohort.[107] The use of fixed cycle lengths is avoided and attributes such as age or severity of disease, which impact on probabilities of experiencing healthcare events, can be assigned to individuals within the model. This allows greater flexibility when the decision question is complex, however, more time and expertise is generally required to implement such models.[108] Microsimulation can refer to state-transition microsimulation, where the model has a set of mutually exclusive health states[109], or pure discrete event simulation where the use of health states is avoided and individuals instead move from one clinical event to the next.[110]

2.5. Calibration and Validation

One means of proving that a model is useful, and capable of giving reliable estimates of potential consequences of introducing an intervention, is the process of validation and calibration. Validation simply involves comparing model outputs to observed data and assessing the level of agreement. Outcomes such as the predicted cancer incidence rate under particular screening and testing scenarios can be compared with the results of large studies or trials, for example. It is a key method of assessing the suitability of a model if it can be shown that its predictions align with other data sources describing the model outputs. Calibration also involves comparing model predictions with observed data but goes further to adjust the parameters of the model to more accurately predict the empirical data.[111 112]

2.5.1. Need for calibration of screening models

Models used to compare screening interventions generally represent the underlying natural history or course of the disease in question and how screening impacts on this. Natural history parameters such as tumour size can be measured and observed however others such as time to disease onset can only be estimated indirectly. The use of calibration in screening models allows the estimation of parameter values which are not directly observable.[113] This is achieved by running alternative sets of input parameter values through the model to identify those which achieve a predicted output close to the calibration target data.[12]

2.5.2. Calibration methods

The first step in model calibration is deciding which parameters should be calibrated. This is normally restricted to unobservable parameters but it is possible to calibrate all unobservable and observable parameters in one process if the model requires it. [12 114 115] The second is deciding which empirical data to calibrate to. The target data should have a large sample size and be representative of the population included in the model. In screening models, it is important to consider any background screening or testing that may have impacted on the calibration data, so sufficient information needs to be available on the population included in the data to determine this. Examples of calibration target data include population statistics, epidemiologic studies, and registry data.[116]

When performing the calibration it is necessary to judge how close the model predictions are to the target data. This can be done visually or through the use of statistical tests such as least squares, weighted least squares or the likelihood.[117] It is also necessary to choose a search strategy or algorithm to search for the best-fitting parameter values. Potential strategies including the grid search, latin hypercube and computer optimisation methods such as Nelder-Mead[118] and Metropolis-Hastings,[111 119] but many different algorithms exist.[120] Once the best-fitting sets have been identified they can be integrated into the decision model.[121]

Model fitting is a distinct concept to model calibration which is also used in health economics. Model fitting or estimation involves estimating the model parameters from observed data using statistical techniques such as maximum likelihood estimation. Model fitting is often used when the model is relatively simple and the parameters are directly related to the observed data.[122]

Bayesian or multiparameter evidence synthesis is an alternative approach to model calibration that uses Markov chain Monte Carlo to draw values from the joint posterior distribution of the parameters, based on prior information on both the parameters and the calibration targets and the likelihood of the different sets of parameter values.[123 124] Most

applications of this approach have been in relatively simple models as it requires complex computations and can be time consuming.[114 122 124]

2.6. Types of uncertainty in decision modelling

No model can be a perfect representation of reality and choices are made at every step of model development, including deciding on health states and choosing input data, that introduce uncertainty. The key types of uncertainty in economic modelling are parameter, structural and methodological uncertainty.

2.6.1. Parameter uncertainty

Normally all data parameters in a model are subject to some degree of uncertainty. Parameter uncertainty relates to the sample size informing each estimate and variance in the data used to estimate the parameter. It also relates to differences in how individuals respond to the effects of a disease or an intervention and the choice of data used to inform the model, where various options are often available. Methods to deal with parameter uncertainty include deterministic and probabilistic sensitivity analysis. In deterministic sensitivity analysis, parameter values are varied one-by-one to test the impact on the model's results. A probabilistic sensitivity analysis varies all parameters at once, with parameters being sampled from their respective distributions (rather than simply using the mean parameter values).[122]

2.6.2. Structural uncertainty

When developing a model, assumptions need to be made about its structure. Examples of such structural assumptions include types of health states and adverse events included, when transitions between health states are expected to occur, and the duration of treatment effects. Choice of relevant comparators may also fall under the term 'structural uncertainty'.[125] Methods to deal with structural uncertainty include scenario analyses and model averaging, where the results from models with different structural assumptions are averaged and weighted by some measure of their credibility.[125]

2.6.3. Methodological uncertainty

Choices must also be made when choosing analytic methods in an economic evaluation including the perspective of the evaluation, whether costs and outcomes occurring in the future should be discounted to current values, and by how much, and what measures to use to estimate costs and health outcomes. Methodological uncertainty has been addressed in many countries through the use of a 'reference case'. This is a template or assumptions to be used for all economic evaluations which is dictated by a governing body. In the UK the reference case is provided by NICE.[89 126] Deviations from the reference case must then be justified and explained. Methodological uncertainty can also be addressed through the use of sensitivity analysis.

2.6.4. Value of Information analysis

Value of information analysis is one means of quantifying uncertainty in decision modelling. It considers the question of whether the cost of acquiring additional information to reduce uncertainty i.e. the cost of conducting additional research, is worth the increased certainty that would result in terms of decision making.[127] Value of Information methods can be used to identify particular areas of uncertainty on which to focus future research studies e.g. cost or utility parameters. They can also be used to prioritise projects in terms of the expected return on investment, when different projects are competing for research funds.[128]

2.7. Overview of challenges in cost-effectiveness analyses of screening interventions

2.7.1. Parameter uncertainty

Cost-effectiveness models of screening interventions are particularly susceptible to parameter uncertainty as they involve not only the screening test but also the subsequent diagnosis, treatment and after-care of the disease. As mentioned, trials reporting all relevant outcomes for a screening model are therefore rare. To accurately model the impact

of a screening test, data must be identified on the prevalence of the disease in the population, ideally by age-group, severity and other modifying factors, and the accuracy of the test in these populations. Data must also be available on the accuracy of any subsequent tests such as biopsy in each of these groups, and dependent on the initial test. As the NICE Diagnostic Assessment Programme Manual states, the amount and quality of the evidence directly relating to diagnostic tests is generally much lower than for other technologies such as drugs.[129] Any uncertainty in such parameters must therefore be fully reflected in the model. Other parameters subject to uncertainty in screening models include the impact on quality of life and resource use associated with testing and treatment and both unobserved and observed natural history parameters.

2.7.2. Uncertainty in natural history of disease

As discussed, it is often the case in screening models that the natural history of the condition in question is not clearly defined, particularly when natural history parameters are not directly observable. This may lead to many different possible means of characterizing the health states in a model.

2.7.3. Uncertainty as to relevant screening interventions

A further complicating factor in cost-effectiveness models of screening interventions is that the frequency of screening, population to screen, diagnostic test or tests used, and their order and combination are all subject to changing practice, with rapid technology advances common. This makes the identification of relevant comparators challenging. These issues will be addressed further in the subsequent chapters.

CHAPTER 3. SYSTEMATIC REVIEW OF COST-EFFECTIVENESS MODELS IN PROSTATE CANCER: EXPLORING NEW DEVELOPMENTS IN TESTING AND DIAGNOSIS

3.1. Introduction

As discussed in sections 1.1 and 2.7, screening interventions are particularly difficult to analyse as they involve not only the screening test but also the subsequent diagnosis, treatment and after-care of the disease (Figure 1). For this reason, a systematic review of the literature to identify recent cost-effectiveness analyses and the tests and diagnostic methods compared, the treatments considered, and how the natural history of the disease has been modelled, is a useful starting point for any new analysis. This can also help to identify areas of parameter and structural uncertainty if different models have used different evidence sources and structures.

As innovations that aim to address the overdiagnosis associated with prostate cancer screening become available, healthcare policy makers must make informed decisions regarding their use in national screening strategies. As such, it is essential to establish the cost effectiveness of these developments, and their combinations, to make rational decisions about the allocation of limited healthcare resources. This systematic review aimed to identify published cost-effectiveness models assessing the impact of novel innovations on the costs and outcomes of prostate cancer diagnosis.

In terms of a PICOS[130], The population of interest (P) was men at risk of developing prostate cancer and the interventions reviewed (I) were novel biomarkers and MRI-guided biopsy techniques as prostate cancer diagnostic tools. The alternatives against which the interventions were compared (C) were standard diagnostic tools such as the PSA test, TRUS-guided biopsy, or no intervention, and the outcome considered (O) was the cost-effectiveness of these interventions in comparison with each other. The study design (S) was model-based economic evaluations of screening or diagnostic strategies including cost-effectiveness, cost-utility, cost-consequence and cost-benefit analysis.

This review also determines the current evidence base and provides an overview of model characteristics. It provides information on novel tests, how they have been modelled, data available to populate such models and general conclusions on cost-effectiveness. It assesses the limitations of available models and has, in turn, assisted in the adaptation of a cost-effectiveness model to a new setting (Chapter 5).

The work described in this chapter has been published as a manuscript in Value in Health.[131]

3.2. Methods

3.2.1. Study Selection

Study selection proceeded from title and abstract screening against the eligibility criteria (section 3.2.3) through full-text review to data extraction. A second reviewer (Josie Morley, University of Bristol) independently screened 10% of the titles and abstracts and performed data extraction on 20% of the included studies. Studies were categorized according to cost-effectiveness analyses of new (1) biomarkers/tests/risk models for screening in prostate cancer, (2) biopsy methods for definitive diagnosis after an initial triage screening test in prostate cancer, and (3) follow-up testing and diagnostic strategies for men initially found to have no prostate cancer.

3.2.2. Search Strategy

In April 2021, studies were identified by searching the NHS Economic Evaluation Database (EED) (2009-2014), Medline, EMBASE, HTA databases, NICE guidelines, UK National Screening Committee guidance, and reference lists from relevant studies. The review was restricted to evidence from January 2009 onward to reflect current practice in screening and testing for prostate cancer and because the aim was to identify novel tests in prostate cancer diagnosis. Search terms included free text and MESH terms (Appendix 2). The search was limited to English language publications. Conference abstracts were excluded.

3.2.3. Eligibility Criteria

As specified in the PICOS, studies were included if they met the following criteria:

- Model-based (rather than trail-based) economic evaluation of screening or diagnostic strategies for prostate cancer beyond the standard PSA test plus TRUS-biopsy
- Cost-effectiveness, cost-utility analysis, cost-consequence analysis and cost-benefit analysis
- Any test for diagnosing or ruling out prostate cancer
- Any subsequent follow-up regime (aside from PSA testing) when prostate cancer has not been identified at initial biopsy
- Any country or type of health system

3.2.4. Data extraction

Data extraction forms were developed and pilot-tested on a random sample (5%) of included studies and refined accordingly. The data extraction form is shown in Appendix 3. Information was extracted from each included study on:

- Context (perspective and country)
- Characteristics of the strategies compared (e.g., frequency of testing and threshold for a positive result)
- Population strategy applied to (i.e. screening start and stop age and the prevalence of prostate cancer)

- Type of outcome measure (including cost per Quality Adjusted Life Years (QALY) gained and life-years gained)
- Cost-effectiveness result
- Characteristics of the model including model type (e.g. decision tree, Markov model), structure (how clinical pathways are represented), and handling of disease natural history
- Sensitivity analyses (including the extent to which uncertainty in the cost-effectiveness result had been quantified)
- Evidence sources for quality of life, resource use and adverse effects
- Evidence sources for accuracy of tests

3.2.5. Quality assessment

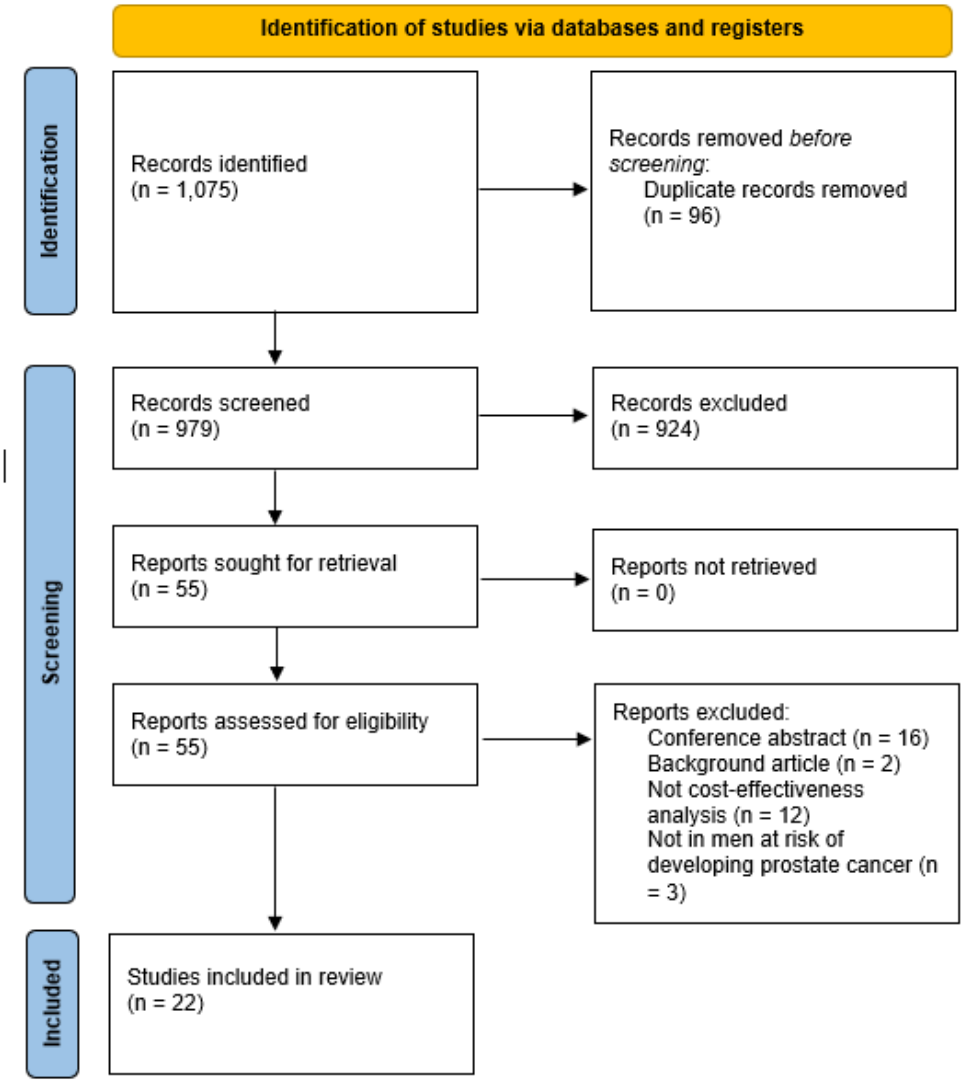
The purpose of the review was to determine the current evidence base and provide an overview of the characteristics of available models. Therefore, a formal quality checklist was not used to exclude studies from the review. Nevertheless, existing economic evaluation checklists were used as a guide to reporting the studies.[12 132] The quality of the included economic evaluations was assessed using the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.[133] A score of 0, 1, or 2 was allocated for each criterion corresponding to a decision of criterion not met, criterion met, or criterion not applicable. Risk of bias was assessed using the Bias in Economic Evaluation (ECOBIAS) checklist.[134] Every item was rated as yes, no, partly, unclear, or not applicable. The review follows published reporting standards for reviews of economic evaluations.[135-137]

3.3. Results

In total, 1075 studies were identified. Most studies were excluded at the abstract stage because they were not model-based economic evaluations or did not compare tests for diagnosing prostate cancer. After removing duplicates and checking for eligibility, 55 full-text articles were retrieved (Figure 2). [138] Of the 55 full-text articles, 22 studies were included in the review. A total of 16 articles were excluded because these were conference abstracts and the rest were excluded because they (1) were background articles rather than

original studies, (2) were not cost-effectiveness analyses, or (3) did not consider a population of men at risk of developing prostate cancer.

Figure 2. PRISMA 2020 flow diagram to demonstrate studies included and excluded from the review



The remainder of this chapter will describe the included studies starting with a breakdown of those that compared new biomarker tests, new biopsy methods, or follow-up strategies for men with a previous negative biopsy, and the tests and methods compared. Model inputs will then be discussed including data on test accuracy, quality of life and resource use. This is followed by a discussion of model characteristics including model type, model structure, sensitivity analyses explored, mechanism of screening benefit assumed and reporting of overdiagnosis. A breakdown of cost-effectiveness results by biomarker tests,

biopsy methods and follow-up strategies will then be presented, followed by results of sensitivity analyses and an assessment of the quality of the included studies.

3.3.1. Study type

Of the 22 studies, eleven compared the cost-effectiveness of new urinary or blood biomarkers to each other or to the standard of care (a PSA test alone). Another eight studies compared different approaches to prostate biopsy and three studies compared follow-up strategies in men who have a negative initial biopsy result (Table 3). The studies were based in the US (n = 6), UK (n =6), Netherlands (n = 4), Hong Kong (n = 1), Germany (n = 1), China (n=1), Sweden (n=1), and Canada (n = 1). One study compared results for France, Germany, Spain and Italy.[139] All but three studies carried out a cost-utility analysis where outcomes were measured in QALYs gained. [140-142] The other three were cost-consequence analyses reporting the number of tests and biopsies carried out and expected overall diagnostic costs. [141] [140]

Table 3. Studies included following full-text screening

Author	Year	Country	Patient population	Age	Assumed prevalence of PCa*	Strategies compared
Strategies compared - biomarkers						
Karlsson et al [143]	2021	Sweden	All men	55-69	NR	1. no screening 2. quadrennial screening for men aged 55-69 years with PSA test alone 3. quadrennial screening for men aged 55-69 with PSA test and reflex Stockholm3 test for PSA values above 1, 1.5 and 2 ng/mL, respectively
Kim et al [142]	2020	UK	referred from primary care for elevated PSA	66	NR	1. mpMRI and biopsy all 2. mpMRI all and biopsy if positive 3. mpMRI all and biopsy if PSA density ≥ 0.15 4. mpMRI all and biopsy if PSA density ≥ 0.1 5. phi all and mpMRI and biopsy if phi ≥ 25 6. phi all and mpMRI and biopsy if phi ≥ 30
Teoh et al [144]	2020	China	Patients with normal DRE undergoing opportunistic PSA testing	50-75	NR	1. Biopsy if PSA 4-10 ng/ml 2. Biopsy only if PSA 4-10 ng/ml and PHI > 35
Bouttell et al[140]	2019	Hong Kong	Normal DRE, PSA 4-10 ng/ml	NR	10.9%	1. Biopsy all 2. Biopsy only if PHI > 25 3. Biopsy only if PHI > 35 4. Biopsy only if PHI > 55
Govers et al [139]	2019	France, Germany, Italy and Spain	Men who under current guideline concordant management, would undergo initial TRUS biopsy	NR	France – 47% Germany – 49% Italy – 37% Spain – 33%	1. Biopsy all 2. Biopsy only if SelectMDx +
Govers et al [145]	2018	US	elevated PSA or abnormal DRE	NR	46.4%	1. Biopsy all 2. Biopsy only if SelectMDx +

Sathianathen al[146]	2018	US	PSA > 3 ng/ml	50	29%	1. Biopsy all 2. Biopsy only if SelectMDx + 3. Biopsy only if PHI + 4. Biopsy only if EPI + 5. Biopsy only if 4Kscore +
Dijkstra et al[147]	2017	Holland	PSA > 3 ng/ml	NR	44.4%	1. Biopsy all 2. Biopsy only if SelectMDx +
Heijnsdijk et al[148]	2016	Holland	PSA > 3 ng/ml	50-75	NR	1. Biopsy all 2. Biopsy only if PHI > 25
Schiffer et al [141]	2012	Germany	PSA > 4 and/or suspicious DRE in a urological outpatient centre setting	66	24%	1. Biopsy all 2. Biopsy only if UPA-PC +
Nichol et al [149]	2011	US	PSA 2-10 ng/ml	50-75	25%	1. Biopsy all 2. Biopsy only if PHI +
			PSA 4-10 ng/ml	50-75	25%	
			PHI+ at PSA 2-10 ng/ml	50-75	29.6%	
			PHI+ at PSA 4-10 ng/ml	50-75	30.3%	
			PSA > 10 ng/ml	50-75	66.70%	
Strategies compared - biopsy methods						
Callender et al [28]	2021	UK	All men	55-69	NR	1. No screening 2. Age-based screening with biopsy if PSA ≥ 3 3. Age-based screening with MRI if PSA ≥ 3 and biopsy if abnormal findings 4. Risk-stratified screening with biopsy if PSA ≥ 3 5. Risk-stratified screening with MRI if PSA ≥ 3 and biopsy if abnormal findings
Barnett et al [150]	2019	US	Biopsy-naive men with PSA > 4 ng/ml	55-69	NR	1. Standard biopsy for all 2. mpMRI, if positive combined biopsy 3. hybrid ^{18}F-choline PET/mpMRI, if positve combined biopsy For 2&3 additional strategies of using Likert or PI-RADSv2 scores to determine positive results and no further biopsy or additional standard biopsy if negative

Barnett et al [151]	2018	US	Biopsy naive men with PSA >4 ng/ml	55-69	NR	1. Standard biopsy for all 2. MRI, if positive targeted fusion biopsy 3. MRI, if positive combined biopsy For 2&3 additional strategies of no further biopsy or additional standard biopsy if negative
Faria et al [152]	2018	UK	men at risk of PCa referred to secondary care for further investigation	NR	38%	383 clinically feasible combinations of MPMRI, TRUSB, and TPMB, in addition to the use of TRUSB and TPMB in isolation
Pahwa et al [153]	2017	US	biopsy-naive men recommended for prostate biopsy on basis of abnormal DRE or elevated PSA	41-50	37%	1. Standard biopsy for all 2. MRI + cognitively guided biopsy 3. MR imaging/US fusion biopsy 4. in-gantry MR imaging-guided biopsy For 2-4 additional strategies of no further biopsy or additional standard biopsy if negative
				51-60	44%	
				41-70	50%	
				61-70	65%	
Venderink et al [154]	2017	Holland	biopsy-naïve men with elevated PSA or abnormal DRE	NR	25%	1. TRUS-guided biopsy for all 2. mpMRI, if suspicious MRI-TRUS fusion–guided biopsy 3. direct in-bore MRI-guided biopsy
Cerantola et al [155]	2016	Canada	biopsy-naive men with clinical suspicion of PCa (based on DRE and PSA values 4-10 mg/) with life expectancy of 20 y	60-65	24%	1. TRUS-guided biopsy for all 2. MRI-targeted biopsy
de Rooij et al[156]	2013	Holland	elevated PSA level (>4 ng/ml)	60	25%	1. TRUS-guided biopsy for all 2. MRI-guided biopsy
Strategies compared - follow-up strategies in men with negative biopsies						
NICE guideline[157]	2019	UK	raised PSA, negative MRI and/or negative prostate biopsy	66-75	58.2%	Different follow-up strategies, including screening test (PSA density, velocity, doubling time, % free forms) PCA3 or PHI, at different frequencies and different thresholds for triggering further investigation. Diagnostic stage possibly including MRI techniques.
Nicholson et al [158]	2015	UK	men referred for second biopsy because, following negative initial biopsy result, clinicians still suspect malignant PCa present	NR	24%	1. clinical assessment 2. clinical assessment + PCA3 3. clinical assessment + phi 4. clinical assessment + PCA3 + phi 5. clinical assessment + mpMRI 6. clinical assessment + mpMRI + PCA3 7. clinical assessment + mpMRI + phi

						8. clinical assessment + mpMRI + PCA3 + phi
Mowatt et al[159]	2013	UK	suspected PC with a prior negative/inconclusive biopsy, with indications for repeat biopsy (i.e. sustained suspicion of PC as a result of clinical and/or pathological findings)	60	24%	1. TRUS-guided biopsy for all 2. T2-MRI 3. MRS 4. DCE-MRI 5. T2-MRI or MRS 6. T2-MRI or DCE-MRI

**Estimate of prevalence used in analysis. Values came from literature or were measured in the primary study cohorts*

Legend: PHI – Prostate Health Index, PSA – Prostate-Specific Antigen, mpMRI – Multiparametric Magnetic Resonance Imaging, EPI - ExoDx® Prostate(IntelliScore), TRUS – Transrectal Ultrasound, MRI – Magnetic Resonance Imaging, MRS - Magnetic Resonance Spectroscopy, DCE-MRI - Dynamic contrast-enhanced magnetic resonance imaging, DW-MRI - Diffusion-weighted magnetic resonance imaging, UPA-PC - Urinary Proteome Analysis for PCa diagnosis

Strategies compared – biomarkers

Of the eleven studies which compared different biomarkers (defined as laboratory measurements which have both diagnostic and prognostic utility[160]), five compared PSA-based testing to PSA plus Prostate Health Index (PHI) testing (described in Appendix 1).[140 142 144 148 149] Four of the five studies compared strategies where PHI was introduced as an additional (reflex) test in men with elevated PSA before deciding who to refer for biopsy. Only Kim et al considered a strategy where PHI was considered a replacement for the PSA test.[142] The PHI cut-off used to trigger further investigation was 25 in Heijnsdijk et al[148] and Nichol et al[149] and 35 in Teoh et al.[144] Bouttell et al[140] and Kim et al compared cut-offs of 25, 35 and 55 and 25 and 30, respectively. Three studies compared PSA-based testing to PSA plus SelectMDx testing.[139 145 147] Similar to PHI, all three studies compared strategies where all men with an elevated PSA are biopsied to those where only men with an elevated PSA and a positive score on the SelectMDx are biopsied.

Of the final three studies, Schiffer et al compared introducing the Urinary Proteome Analysis for PCa diagnosis (UPA-PC) test either before first or before re-biopsy, and submitting only patients with a positive UPA-PC test result to prostate biopsy, to the standard diagnostic routine of biopsy alone.[141] Sathianathen et al compared PHI at a cut-off of 25, 4Kscore at a cut-off of 7.5%, EPI at a cut-off of 15.6 and SelectMDx at a cut-off of -2.8 to each other and to a simple PSA test at a cut-off of 3 ng/ml to triage men for TRUS-guided biopsy.[146] Karlsson et al compared screening with a PSA test alone to screening with a reflex Stockholm3 test for PSA values above 1, 1.5 and 2 ng/mL, respectively.

Of the eleven studies comparing different biomarkers, six referred to TRUS-guided biopsy to confirm diagnosis and one did not report the biopsy method assumed.[148] Only one study also reported results from an alternative scenario of undergoing mpMRI and, if positive, MRI-guided biopsy as opposed to TRUS guided.[146] Only two studies, Nichol et al[149] and Heijndisk et al[148], modelled repeat testing in the context of screening, assuming annual and 4-yearly screening respectively. Nichol et al based their assumption of annual screening on the 2009 American Urological Association (AUA) recommendations[161] and Heijndisk et al

modelled 4-year intervals according to the ERSPC protocol.[21] The other five studies did not model repeat testing.

Strategies compared - biopsy methods

Seven studies compared TRUS-guided biopsy to other biopsy methods in men with a suspicion of prostate cancer indicated by a PSA test. The different biopsy methods compared included MRI-targeted methods (fusion, combined, cognitively-guided, and in-gantry/in-bore) and template mapping biopsy. The definitions of these biopsy methods are given in Appendix 1. Three of the studies comparing biopsy methods [22 28 29] modelled repeat screening, assuming that men would be screened every 2 years based on the 2013 American Urological Association (AUA) guideline[162] or every 4 years based on the ERSPC protocol.[21]

Five of these studies compared standard biopsy to one or more MRI-targeted methods with the MRI acting as a triage test to reduce the number of biopsies carried out.[151 153-156] In 2 of these 5 studies [151 153] strategies where standard biopsies were offered to men with negative MRI results were also compared. Faria et al compared 383 clinically feasible combinations of mpMRI, TRUS guided biopsy and template prostate mapping biopsy (TPMB), in addition to the use of TRUS guided biopsy and TPMB in isolation.[152] These included strategies using mpMRI to decide whether a biopsy is necessary and to target the biopsy, and strategies giving a TRUS-guided biopsy and then using mpMRI to decide whether a repeat biopsy is warranted. The final study, Mowatt et al[159], compared Magnetic resonance spectroscopy imaging (MRSI), Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), Diffusion-weighted magnetic resonance imaging (DW-MRI), and standard T2-MRI to identify areas of the prostate for targeting in a subsequent TRUS guided biopsy to standard TRUS for guiding biopsy.

Strategies compared – follow-up strategies in men with negative biopsies

Three studies [157 158 163] compared follow-up strategies for men with raised PSA, negative MRI and/or negative prostate biopsy. The health economic report for the recent NICE guideline on managing prostate cancer compared 191 follow-up strategies for this group.[157] A follow-up strategy consisted of the frequency of testing (3-monthly, 6-monthly, 1 yearly, 2 yearly, 3

yearly), type of screening test (PSA, PSA velocity, PSA density, % free PSA, PSA doubling time, PSA density in transition zone, PCA3, PHI) and the threshold at which cases were identified as positive. A diagnostic stage involving an option of imaging the prostate using MRI techniques was also included. Both TRUS and template mapping biopsy were considered.

The HTA report by Nicholson et al [158] compared clinical assessment on its own to clinical assessment + PCA3, PHI, PCA3 + PHI, mpMRI, mpMRI + PCA3, mpMRI + PHI or mpMRI + PCA3 + PHI as reflex tests for men who have been referred for a second biopsy. Two types of prostate biopsies were included - TRUS and transperineal, with the option of mpMRI informing which type is carried out. The final study, Mowatt et al[161]assessed the cost-effectiveness of magnetic resonance spectroscopy (MRS), dynamic contrast-enhanced MRI (DCE-MRI) and diffusion-weighted MRI (DW-MRI) in men with a previous negative biopsy. The analysis evaluated the use of MRS and MRI in the context of using it to identify areas of the prostate for targeting in a subsequent TRUS biopsy.

Patient population

In terms of age, the youngest cohort modelled was in Pahwa et al[153] who compared results for a cohort of men aged 41-50. The NICE guideline stated that in the studies used to inform their model, the average age was between 62 and 73 years old and. A baseline age of 66 was therefore thought to be appropriate. They also stated that the committee advised the age of 75 to be a realistic age for screening to stop as it is unlikely that men diagnosed after this point would be considered for radical therapy. Half the studies did not report the age of the cohort modelled.

The assumed prevalence of prostate cancer in the cohorts modelled, which was either estimated from the literature or measured in the primary study cohorts, varied from 10.9% for Chinese men with a normal DRE and PSA level of 4-10 ng/ml [140] to 66.7% in US men with a PSA level > 10 ng/ml.[149] A prevalence of between 24% and 30% was commonly used. Several studies also reported the percentage of prostate cancers that were assumed to be

high grade or significant. Dijkstra et al[147] assumed 51.2% of all prostate cancers to be high grade[164], Govers et al assumed 49.1%[145], Pahwa et al assumed 50% and Bouttell et al 26%.[165]

Treatment types

A total of ten studies reported the percentage of patients with high grade/clinically significant or low grade/clinically insignificant prostate cancer allocated to each type of treatment (Table 4). The percentage of high grade patients allocated to a radical treatment (prostatectomy, radiotherapy, brachytherapy, hormone therapy or ADT) varied from 100% [151] to 65%.[156] The percentage of low grade patients allocated to a radical treatment varied from 90%[153] to 20%.[147 156] Of the other twelve studies, five did not include treatment in their timeframe [140-142 158 166], six stated that individual treatments were modelled but did not give the allocation ratio [28 143 146 148 152 159], and one stated that they did not explicitly model individual treatments.[149]

Table 4. Treatment allocation assumed (%)

Study	Dijkstra et al[147]	Govers et al[145]	Barnett et al[151]	Pahwa et al[153]	Venderink et al[154]	Cerantola et al[155]	de Rooij et al[156]	Barnett et al[150]	Govers et al [139] Spain	Govers et al [139] Italy	Govers et al [139] Germany	Govers et al [139] France	NICE guideline intermediate risk (high risk)[157]
High Grade/Clinically significant													
RP	70	54	100	32	70	30	40	100	34	56	67	58	16 (12)
RT	25	40	-	18	25	30	25	-	36	19	18	15	35 (35)
BY	-	-	-	8	-	-	-	-	5				3 (1)
BY+EBRT	-	-	-	-	-	10	-	-	-				-
ADT	-	-	-	33	-	-	-	-	21	19	10	24	-
RT+ADT	-	-	-	-	-	30	-	-	-				-
HT	-	-	-	-	-	-	-	-	-				22 (48)
WW	5	6	-	2	-	-	18	-	4	6	5	4	-
AS	-	-	-	2	5	-	18	-	-				25 (5)
Low Grade/Clinically insignificant													
RP	10	50	50	57	40	35	10	50	49	65	50	34	18
RT	10	30	-	7	10	35	-	-	19	11	16	9	20
BY	-	-	-	16	-	15	10	-	17				7
ADT	-	-	-	8	-	-	-	-	5	8	5	16	-
HT	-	-	-	-	-	-	-	-					9
WW	-	-	-	5	-	-	40	-					-
AS	80	20	50	5	50	15	40	50	9	16	29	41	47
Source	[156] expert opinion	[167]	[168]	[169]	expert opinion	[159]	[169], [170], expert opinion	[171]	[172]	[173]	[174]	[175] [176]	[177]

Legend: RP – Radical Prostatectomy, RT – Radiotherapy, BY – Brachytherapy, EBRT – External Beam Radiotherapy, ADT – Androgen Deprivation Therapy, WW – Watchful Waiting, AS – Active Surveillance, HT – Hormone Therapy

3.3.2. Model inputs

Accuracy data

All but six studies [28 143 144 149 155 158] explicitly reported the sensitivity and specificity of the tests compared. The assumed sensitivity of a standard biopsy ranged from 0.9 based on ERSPC data [148 178] to 0.46 based on de Rooij et al [146 153 156]. The biomarkers were generally assumed to be either particularly sensitive or particularly specific. PHI at a threshold of 20, for example, had the highest reported sensitivity (1, but at a very low specificity of 0.08).[142] PHI at a threshold of 55 had the highest reported specificity (0.974, but at sensitivity of 0.129).[140]

The MRI-targeted biopsy methods generally had a better balance of sensitivity and specificity, ranging from a sensitivity of 0.965 (specificity 0.597) for MRI using a Prostate Imaging–Reporting and Data System (PI-RADS) threshold ≥ 3 to [151 179] to 0.770 (specificity 0.68) using fusion biopsy to detect clinically significant disease.[146 154 180] Appendix 4 details the accuracy estimated used along with their evidence sources.

Quality of Life

As detailed in Table 5, all but 3 studies[140-142] assigned utility scores to various aspects associated with testing including screening attendance, the biopsy procedure, diagnosis of cancer, treatment, active surveillance, advanced or metastatic cancer, post-treatment or recovery, adverse events associated with biopsy and treatment, and palliative therapy. Where utility estimates were sourced from studies directly measuring health related quality of life, the most common methods used were standard gamble[181-183] and time-trade off [184 185] and the most common instruments used were EQ-5D [85 184-187]and SF-12.[188 189]

Nine studies [28 139 145 147 148 151 154 158 190] sourced all utility estimates used in their model from Heijnsdijk et al[190] who in turn obtained their utility estimates from the Cost-Effectiveness Analysis Registry and various additional studies.[181 186 191-201] These studies were from

different countries and settings and used mixed evaluation techniques. The other studies sourced their utility estimates from various unrelated publications, also in different countries and settings. Although this approach indicates that there is likely no alternative common source for these utility parameters, this is against best practice as the values cannot be considered to be equivalent when measured in different populations.[202]

Ara et al recommend that when health state utility values are sourced from the literature details should be given on searches, inclusion/exclusion criteria, and the quality and relevance of included studies.[203] They also state that a justification should be given for the utility values chosen. None of the included studies provided this level of detail making it hard to establish whether the estimates used are relevant. The majority of studies that assigned a disutility to the biopsy procedure, for example, cited Heijnsdijk et al[190] as the source. This study reported a utility decrement of 0.1 that lasted 3 weeks following a biopsy, equivalent to 18.9 days spent in perfect health. This utility value, however, was taken from an earlier study[191] that focused on breast cancer biopsy and the duration of decrement was an assumption based on clinical opinion. No reference was made in any study to an attempt to identify a disutility value associated specifically with prostate biopsy.

Only five studies fully reported the uncertainty in the disutility estimates used[28 150 151 153 156] , suggesting that this uncertainty was not accounted for in the other studies. As QALY estimates can often have a substantial impact on the intervention considered most cost-effective, it is important that any underlying uncertainty in these estimates is fully accounted for.

Table 5. Disutility estimates used for prostate cancer states, tests and treatments in the identified cost-effectiveness models (annual values)

Study	Biopsy	Diagnosis	RP	RT	AS	Advanced cancer	Post-treatment	AEs	Other	Source	Unit used for uncertainty
Barnett et al 2018[151] Barnett et al 2019[150]	0.006 (0.00346–0.0075)	0.017 (0.0125–0.0208)	0.247 (0.0917–0.323)	-	0.03 (0.0–0.15)	0.3 (0.3–0.38)	0.05 (0.0–0.07)	0.0161 (0.00969–0.0291) (post-biopsy infection)	0.0002 (0.0–0.00019) (PSA screening) 0.00077 (0.00038–0.0012) (MRI) 0.60 (0.14–0.76) (Palliative therapy)	[190] [204]	Range
Cerantola et al 2016[155]	-	-	-	-	-	-	0.08	-	0.22 (relapse)	[205]	No uncertainty
de Rooij et al 2014[156]	-	-	0.33 (0.29)	0.27 (0.30)	0.16 (0.19)	-	-	-	-	[181]	Standard deviation
Dijkstra et al 2017[147]	0.006	0.017	0.228	0.247	0.03	-	0.05	-	-	[190]	No uncertainty
Faria et a 2018[152]	0.007 (0.006 to 0.008) (TPM biopsy)	-	-	-	-	0.137	-	-	-	[186], PROMIS IPD [25], [187]	Not reported
Govers et al 2018[145] Govers et al 2019[139]	0.006	0.017	0.228	0.247	0.03	-	0.05	-	-	[190]	No uncertainty
Heijnsdijk et al 2016[148]	0.006	0.017	0.247	0.228	0.03	0.3	0.05	-	0.0002 (Screening attendance) 0.60 (Palliative therapy)	[190]	No uncertainty
Mowatt et al 2013[159]	-	-	-	-	-	0.365 (0.04)	-	0.16 (urinary incontinence) 0.17 (bowel problems 0.12 (erectile Dysfunction)	0.11 (0.0133) (Localised (undiagnosed)) Localised (diagnosed) <6 months 0.11 (0.0133) Localised (diagnosed) 6–12 months 0.09 (0.014427) Localised (diagnosed) 12–51 months 0.1 (0.015328) Localised (diagnosed) ≥52 months 0.12 (0.018276) 0.19 (0.014625) (Locally advanced (undiagnosed))	[184] [185] [182]	Standard error of mean. Uncertainty only reported for cancer states

									Locally advanced (diagnosed) <6 months 0.19 (0.0146) Locally advanced (diagnosed) 6–12 months 0.17 (0.0156) Locally advanced (diagnosed) 12–51 months 0.18 (0.0149) Locally advanced (diagnosed) ≥52 months 0.24 (0.0205)		
NICE guideline 2019[157]	0.004, 0.007 (Template mapping biopsy)	-	-	-	-	0.137	-	-	0.027 (low-risk) 0.029 (intermediate-risk) 0.027 (high-risk)	[159 187 188 190 204 206]	No uncertainty
Nichol et al 2012[149]	0.027	-	-	-	-	-	-	-	0.2 (0.3) (PCa)	[207 208 209]	Standard deviation. Uncertainty only for PCa
Nicholson et al 2015[158]	0.006	-	-	-	-	-	-	-	-	[190]	No uncertainty
Pahwa et al 2017[153]	0.027	Only lifetime QALYs reported								[207]	Yes
Sathianathen et al 2018[146]	0.004	-	0.14	-	0.03	0.42	0.05	-	-	[183 190 207]	No uncertainty
Venderink et al 2017 [154]	0.006	0.02	0.25	0.23	0.03	0.55	0.05	-	-	[190]	No uncertainty
Callender et al [28]	-	-	-	-	-	-	-	-	0.07 (0.12 – 0) (PCa)	[190]	95% CI
Teoh et al [144]	0.027	-	-	-	-	-	-	-	0.2 (PCa)	[181 183 208 210]	No uncertainty
Karlsson et al [143]	0.1	0.2	0.33 (part 1), 0.23 (part 2)	0.27 (part 1), 0.22 (part 2)	0.03	0.6	0.05		0.60 (Palliative therapy), 0.01 (PSA test)	[190]	No uncertainty

Legend: RP: Radical Prostatectomy, RT: Radiation Therapy, AS: Active Surveillance, AE: Adverse Events, TPM: Template Mapping, IPD: Individual Patient Data, QALY: Quality-adjusted Life Year, PCA: Prostate cancer

Resource use

The majority of studies took a healthcare perspective for the analysis. Two studies stated that a societal perspective was taken but did report the societal costs that were included.[148] [149] Two studies included productivity costs in terms of missed days of work when a patient undergoes a test or treatment. [143] [153] No study gave a justification for the perspective taken. The main costs included were the cost of testing, biopsy and subsequent management strategy. Thirteen studies included costs of complications arising from biopsy [139] [140] [142] [145] [147] [150-154] [157-159]. Only seven studies explicitly stated that costs associated with complications arising from treatment were included. [139] [145] [147] [152] [154] [157] [159]

3.3.3. Model characteristics

Model type

Table 6 details model characteristics including model type, time horizon, and cycle length. Eight combined decision tree/Markov cohort models were identified. In five of these, the decision tree reflected the short term diagnostic process and the Markov model began at the point where patients moved from diagnosis to treatment.[145 152 156 166] In the others, the decision tree captured both diagnosis and treatment and the Markov model was used for post-treatment states.[146 147 154] Eight cohort Markov models [28 141 149-151 155 157 159], 2 continuous time discrete-event microsimulation models (the MISCAN model[148] and the Prostata model[143]), and four decision tree models [140 142 153 158] were also identified. No study provided a justification for choosing one model type over another.

The decision trees generally used data on disease prevalence and accuracy of the tests to categorise men into true positives, false positives, true negatives and false negatives[140 147 154] with some also incorporating clinical significance of cancer.[145 147 152 153 156] The Markov models captured cancer progression and survival. All but four studies developed a *de novo* model.[28 143 148 151]

Eleven studies had a one-year cycle length[28 139 144 145 147 149-151 154-156], two assumed a cycle length of 3 months[157 159] and one had a cycle length of 6 months.[146] The other studies did not report the cycle length assumed. The only study that reported a justification for the cycle length chosen was the NICE guideline which stated that the guideline development committee confirmed that a cycle length of 3 months is sufficient to reflect possible clinical events a person with prostate cancer may experience.[157]

The timeframe of the models varied from when patients reached the treatment stage [140 141] to their entire lifetime.[145 146 148 149 151-153] Three studies had a timeframe of 18 years[139 147 154] as this was the median follow-up time of survival data for patients with prostate cancer, described in the SPCG-4 study.[211] One compared results using a 5, 10, 15 and 20 year time horizon [155], one used a 10 year timeframe because 'after this period no differences were

expected between the strategies' [156], and one used a 30 year timeframe as 'by this stage the majority of the modelled cohorts were dead and the additional QALYs per cycle had fallen to <0.001'.[159] Three studies used a time horizon of 3 years or less, modelling only up to biopsy, which is unlikely to be long enough to capture the impact of timely and accurate diagnosis of prostate cancer, due to its long term nature.[158] [140] [141]

Table 6. Model characteristics

Study	Model type	Progression modelled	Health states in model	Definition of low-risk cancer	Definition of intermediate risk cancer	Definition of high-risk cancer	Time horizon	Cycle length	DSA	PSA
Dijkstra et al[147]	Decision tree/Markov	No	high grade PCa, low grade PCa, missed PCa	G≤6	-	G ≥ 7	18 years	1 year	Yes	No
Sathianathen et al[146]	Decision tree/Markov	No	NR	-	-	-	Lifetime	6 months	Yes	Yes
Govers et al[145]	Decision tree/Markov	No	high grade PCa, low grade PCa, missed PCa	G≤6	-	G ≥ 7	Lifetime	1 year	Yes	No
Faria et al[152]	Decision tree/Markov	Yes	Progression free, metastatic	PSA < 10, G <6	PSA 10-15 or G7	G>8	Lifetime	Not reported	Yes	Yes
Venderink et al[154]	Decision tree/Markov	No	status after prostatectomy, status after radiotherapy, status after active surveillance	-	-	-	18 years	1 year	Yes	No
de Rooij et al[156]	Decision tree/Markov	No	Alive, dead	G3 + 3 or small-size 3 + 4	-	large tumours with a G3 + 3 or ≥3 + 4	10 years	1 year	Yes	No
NICE guideline[157]	Decision tree/Markov	Yes	Low risk, intermediate, high risk, metastatic	G ≤ 6, PSA ≤ 10	G = 7 or 10≤PSA<20	G ≥ 8 and PSA > 20	Lifetime	3 months	Yes	Yes
Nichol et al[149]	Markov cohort	No	Alive, dead	-	-	-	Lifetime	1 year	Yes	Yes
Schiffer et al[141]	Markov cohort	No	NR	-	-	-	Up to treatment	NR	Yes	Yes
Barnett et al[151]	Markov cohort	Yes	G<7, G=7, G>7, extraprostatic or lymph-node positive	G<7	G=7	G>7	Until death	1 year	Yes	No
Cerantola et al[155]	Markov cohort	No	MRGTB/TRUSGB; follow-up of PCa-naive patients with DRE, PSA,	-	-	-	5, 10, 15, and 20 years	1 year	Yes	No

			and TRUSGB as required; low-risk PCa; intermediate/high-risk PCa; active surveillance; curative-intended treatment; biochemical recurrence after curative treatment; metastatic/castration-resistant PCa							
Mowatt et al[159]	Markov cohort	Yes	localised (T1–T2) (low risk); localised (intermediate risk); localised (high risk); locally advanced (T3); metastatic	G ≤ 6, PSA ≤ 10, ≤T1a	G ≤ 7, PSA ≤ 20, ≤T2b	G > 7, PSA > 20,>T2b	30 years	3 months	Yes	Yes
Pahwa et al[153]	Decision tree	No	-	G ≤ 6	-	G ≥ 7	Until death	-	Yes	No
Nicholson et al[158]	Decision tree	No	-	-	-	-	3 years	-	Yes	Yes
Boutell et al[140]	Decision tree	No	-	-	-	-	Up to biopsy	-	Yes	Yes
Heijnsdijk et al[148]	Microsimulation	Yes	T1 G<7, G=7, G>7; T2 G<7, G=7, G>7; T3+ G<7, G=7, G>7, each state can be local or metastatic	-	-	-	Lifetime	-	Yes	No
Barnett et al [150]	Markov cohort	Yes	G<7, G=7, G>7, extraprostatic or lymph-node positive	G<7	G=7	G>7	Until death	1 year	Yes	No
Callendar et al [28]	Markov cohort	No	Healthy, PCa	-	-	-	Lifetime	1 year	Yes	Yes

Kim et al [142]	Decision tree	No	-	-	-	-	Up to diagnosis	-	Yes	No
Teoh et al [144]	Decision tree/Markov	No	PCa, no PCa	-	-	-	25 years	1 year	Yes	Yes
Karlsson et al [143]	Microsimulation	Yes	T1-T2 G<7, G=7, G>7; T3+ G<7, G=7, G>7; Metastatic G<7, G=7, G>7,	-	-	-	Lifetime	-	Yes	Yes
Govers et al [139]	Decision tree/Markov model	No	Treatment, no treatment, delayed treatment	G≤7		G≥7	18 years	1 year	Yes	No

Legend: PCa: Prostate cancer, G: Gleason, NR: Not reported, PSA: Prostate Specific Antigen, MRGTB: Magnetic Resonance Guided Transperineal Biopsy, TRUSGB: Transrectal Ultrasound Guided Biopsy, DRE: Digtal Rectal Examination

Sensitivity analysis

All studies conducted a deterministic sensitivity analysis where input parameters or sets of parameters were varied to see the impact on results. Only half of the studies (11/22) carried out a probabilistic sensitivity analysis where repeated simulations sampled all parameters from their respective distributions to observe the impact on results.[28 143 144 146 149 152 153 156-159] No study carried out a Value of Information analysis to determine the value of further research in prostate cancer screening. [212]

Model Structure and Data Sources to Inform Progression

Only seven of the included models [143 148 150-152 157 163] took account of how prostate cancer progresses through different health states, and how the introduction of a new test might impact on this, and all of these captured this progression differently. In addition, the definition of clinically significant cancer varied across studies.

Mowatt et al[159] simulated progression through no or undetectable cancer, low risk (G ≤ 6, PSA ≤ 10,≤T1a), intermediate risk (G ≤ 7, PSA ≤ 20, ≤T2b), high risk (G > 7,PSA > 20,>T2b), locally advanced (T3), metastatic cancer, and prostate cancer death using the D'Amico risk classification.[159] Patients with localised and locally advanced disease were modelled to progress towards metastatic disease based on age, tumour risk status, and whether or not their cancer was diagnosed and appropriately treated. Patients with undiagnosed cancers faced a higher risk of progression to metastases (based on progression rates observed for patients under watchful waiting), whereas those detected were modelled to progress at rates observed for patients receiving radical treatments. Their modelled progression risks and relative treatment effects (post diagnosis) were based on SPCG-4 data.[213] The SPCG-4 study randomly assigned 695 men with localized prostate cancer to watchful waiting or radical prostatectomy from October 1989 through February 1999 and collected follow-up data through 2017.[213]

Barnett et al[150 151] simulated progression through no prostate cancer, organ-confined prostate cancer based on Gleason score (<7, 7, >7), and extraprostatic or lymph node-positive cancer. This model also included post-treatment states of no recurrence or possible recurrence depending on whether the disease was organ confined. The transition rates from Gleason score 7 to Gleason score > 7 and from all states to extraprostatic cancer were based on estimates from the ERSPC.[214]

Faria et al[152] divided their population into no cancer, clinically non-significant cancer (PSA < 10 ng/ml and Gleason score <6), intermediate risk cancer (PSA 10-15 ng/ml or Gleason score 7) and high risk cancer (Gleason score >8) following the diagnostic process. From there,

patients could transition through progression-free cancer (localised), metastatic cancer and death. Long-term outcomes were based on the Prostate Cancer Intervention Versus Observation Trial (PIVOT),[215] which compared radical prostatectomy and watchful waiting in men with localised prostate cancer, and the Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy (STAMPEDE) trial which tested the addition of further treatments to androgen-deprivation therapy in patients with advanced or metastatic prostate cancer. [214] Calibration to these data sources was used to estimate the probability of transition between the health states.

Similar to Faria et al, the NICE guideline model simulated progression through no cancer, low risk (Gleason score ≤ 6 and PSA ≤ 10 ng/ml), intermediate risk (Gleason score = 7 or 10≤PSA<20) and high-risk disease (Gleason score ≥ 8 and PSA > 20). The probability of progression in undiagnosed cases was based on the watchful waiting arm of the SPCG-4 study[213], similar to Mowatt et al. Progression in diagnosed cases was calibrated to data from a prognostic modelling study by Gnanapragasam et al[214] on the cumulative incidence of prostate cancer specific death at 10 years for difference prostate cancer risk groups. It was also calibrated to data from STAMPEDE on overall survival for people with metastases,[216] similar to Faria et al. Progression rates accounted for the treatment received by people in each risk group, as reported in the two studies. This was the only study to give a reason for the cancer pathways chosen stating that the factors used "have been shown to predict the risk of recurrence after treatment of localised prostate cancer".[157]

Unlike the other models, the Prostata model used by Karlsson et al[143] allowed for prostate cancer onset and progression to correlate with changes in PSA over time. Progression from localised to advanced disease by T stage (T 1-2, 3-4 and metastatic) and Gleason score (≤6, 7, ≥8) was modelled. The model was calibrated to data from ERSPC as well as age-specific cancer staging observed in the Stockholm PSA and Biopsy Register.[217] Age-specific survival was calibrated to the Swedish National Prostate Cancer Register.[218] Similar to the Prostata model, the MISCAN model used by Heijndisk et al[148] simulated progression through different preclinical states defined in combinations of clinical T-stage (T1, T2 and T3), Gleason grade and metastatic stage (local-regional and distant). However, this model did

not take account of the impact of changes in PSA. The transitions through states were also informed by data from the ERSPC.[219]

Dijkstrka et al[147] and Govers et al[145] did not model cancer progression, only survival from high grade prostate cancer, low grade prostate cancer or missed prostate cancer, the transition probabilities for which were based on the SPCG-4 study. Dijkstrka et al classified cancers with a Gleason score <6 as low grade and ≥7 as high grade.[147] Nichol similarly only modelled time to prostate cancer related death which was derived from a systematic literature review as part of the study conducted by Hayes et al.[220] Venderink et al[154] used a cohort Markov model to represent follow-up after diagnosis and treatment. The health states were status after prostatectomy, status after radiotherapy, status after active surveillance, and death. Survival data was again based on the SPCG-4 study.

De Rooij[156] had two health states in their model, alive and dead. They assumed different survival for significant (large tumours with a Gleason score of 3 + 3 or tumours ≥3 + 4) and insignificant (Gleason score of 3 + 3 or a small-size 3 + 4) tumours based on the data from the PCPT trial.[221] Finally, Cerantola et al[155] used a Markov cohort model with states low risk cancer, intermediate risk cancer, metastatic/castration-resistant cancer, biochemical recurrence after curative treatment, and death from prostate cancer. 1-year probabilities of recurrence after curative treatment or active surveillance were sourced from the literature[222-224] as well as the probabilities of developing metastatic prostate cancer (CRPC).[225 226]

Six of the models did not consider stages or grade of cancer, only presence or absence of cancer [28 144-146 149 154]. Four did not model beyond diagnosis [140-142 158]. A failure to consider the complexity of the disease, including stage or grade of cancer and how cancer progresses in diagnosed and undiagnosed cases, calls into question the reliability of the results. The cost-effectiveness of a new test may be overestimated if the cancers it identifies would never progress to cause symptoms or mortality if not identified. The purpose of screening and testing is to identify cancers at an early stage when they are more amenable to

treatment. If cost-effectiveness models do not differentiate between cancer stages it is difficult to measure the effects of early diagnosis.[132]

Reporting of overdiagnosis and mechanism of screening benefit

As described in section 2.1.3, overdiagnosis and overtreatment due to the identification of cancers that would never cause harm or result in symptoms in a man's lifetime, are key factors to consider when testing men for prostate cancer. Only three studies [28 143 148] provided estimates of the impact of screening on overdiagnosis. Both Heijnsdijk et al and Karlsson et al defined overdiagnosed cancers as additional cancers detected through screening that were not detected in the 'no screening' arm.[143 148] Heijnsdijk et al estimated that 5% fewer overdiagnosed cancers would be detected through the use of PHI compared to PSA and Karlsson et al predicted 15% fewer overdiagnosed cancers through the use of Stockholm3 when PSA values were above 2 ng/ml compared to PSA alone. Callender et al estimated age-specific overdiagnosis by multiplying the number of cases by an equation derived from Pashayan and colleagues, defined as the probability that a PSA-detected case would have taken longer than the remaining lifetime to progress to clinical cancer.[227] They found that MRI-first risk-stratified screening was associated with a 10.4% to 72.6% lower probability of overdiagnosis in screen-detected cases, depending on the 10-year absolute risk thresholds at which individuals were eligible for screening.

In addition, different approaches to measuring the benefit of screening in non-overdiagnosed men are possible and the choice of method may impact on results. Stage-shift screening models assume that the benefit associated with screening is due to a shift to a less advanced stage at diagnosis resulting in improved survival. Cure models assume that if cancers are detected earlier that they can be treated, and that curative treatment has the potential to prevent cancer-specific mortality.[228] Only one study, Heijnsdijk et al, explicitly stated that the assumed mechanism of benefit of screening in their model was as a cure proportion which assumes that a percentage of men are cured due to screening and therefore avoid a death from prostate cancer.[148] de Rooij stated that significant tumours that went undetected were assumed to have lower survival as they would receive treatment when the cancer was at a more advanced disease stage.[156] The other studies did not

consider overdiagnosis nor give any detail on the mechanism of benefit of screening assumed.

3.3.4. Quality of included studies

The overall mean percentage of applicable CHEERS reporting criteria met by each study was calculated at 71%, with a range of 37–100% and a median of 68%. Only one study, Mowatt et al[163], satisfied all applicable reporting criteria (scoring 100%) (Appendix 5, Table 1). Risk of bias was assessed using the ECOBIAS checklist. No studies had bias in terms ordinal ICERs (the use of non-cardinal scales for outcomes).[134] Partial biases related to structural assumptions, type of model used, and data identification occurred in most studies (Appendix 5, Table 2).

3.3.5. Cost-effectiveness results

To aid comparison, all reported costs were inflated to the 2020 price year and converted to US dollars, taking purchasing power parities between countries into account. This was done using the web-based tool developed by the Campbell and Cochrane Economics Methods Group (CCEMG) and the Evidence for Policy and Practice Information and Coordinating Centre (EPPI-Centre).[229] In reality the costs are not comparable as different countries have different healthcare systems, care pathways and negotiated prices. Original costs are therefore also reported.

Biomarkers

Of the eleven studies that compared PSA testing to testing with a new biomarker, six studies found that introducing the new biomarker saved costs and increased QALYs gained [139 144-147 149] (*Table 7*). Three did not measure QALYs gained but found that diagnostic costs were reduced.[140-142] Sathianathen et al[146] found that the PHI was more costly and less effective than the SelectMDx strategy and that the EPI provided the highest QALYs gained with an ICER of $58,404 per QALYs gained. Of the studies that considered progression through stages or grades of cancer, Heijnsdijk et al found that PSA+PHI testing saved costs compared to PSA testing alone and resulted in the same QALYs gained[148] and Karlsson et al estimated an ICER of €5,663 for screening using Stockholm3 when PSA values were above 2 ng/ml compared to PSA alone.[143] This would be considered a low cost per QALY gained in Sweden, where the study was set. The results from all studies were generally driven by a decrease in negative biopsies.

Biopsy methods

Seven of the eight studies that compared MRI guided biopsy strategies to each other and to TRUS biopsy found that at least one MRI guided strategy led to additional QALYs gained for the patient, compared with the standard biopsy strategy, at a cost below the WTP threshold. The only exception was Cerantola et al [155] who found that MRI-guided biopsy decreased costs and increased QALYs gained compared to TRUS guided biopsy. De Rooij et al calculated an ICER of €323 per QALY for MRI-guided biopsy compared to TRUS guided

biopsy. Similarly, Venderink et al calculated an ICER of €1,386 per QALYs gained for fusion-guided biopsy compared with TRUS guided biopsy. Pahwa et al [153] found that in-gantry MR imaging-guided biopsy was the most cost-effective of the options compared, with an ICER of $4,147 compared to the next most cost-effective imaging method, cognitively guided MRI.

Barnett et al[151] calculated an ICER of $23,483 per QALYs gained compared to no screening for combined biopsy for patients with PI-RADS score ≥3 and no biopsy for patients with PI-RADS score <3. Faria et al[152] found that the use of mpMRI first and then up to two MRI-targeted TRUS biopsies was cost effective (ICER = £7,076) in a UK setting. These studies all indicate that MRI guided biopsy strategies would be considered good value for money according to the generally accepted cost-effectiveness thresholds in the respective countries. The increased QALYs gained and reduced costs were generally due to an avoidance of the adverse effects and resource use associated with overdiagnosis.

Follow-up strategies

Two of the studies comparing follow-up strategies in men with a previous negative biopsy did not identify a clear indication of cost-effectiveness for any strategy. Nicholson et al[158] found no strategy to be cost-effective at the £20,000 - £30,000 cost effectiveness threshold recommended by NICE and Mowatt et al found the base-case ICER for T2-MRI to be below the UK willingness to pay threshold (£30,000 per QALYs gained) for all cohorts modelled. The NICE guideline[157] concluded that measures derived from PSA tests, including velocity at a threshold of 0.75 ng/ml/year, density at a threshold of 0.15 ng/ml/ml and the percentage of free PSA at a threshold of 15% appear to be the best indicators to trigger further diagnostics within the majority of subpopulations.

Table 7. Cost-effectiveness results from studies

Author	Tests compared	Difference in costs¥	Difference in QALYs*	ICER (cost per QALY gained)∞	Probability cost-effective
Bouttell et al[140]	PHI v PSA	-HK$5,500 (-$943)	NA as study was cost-consequence analysis	NA	NR
Heijnsdijk et al[148]	PHI v PSA	-€33 (-$47)	0	NA	NR
Nichol et al [149]	PHI v PSA	-$201 to -$1,199 (-$243 to -$1,447)	0.01 to 0.08	Dominates	77% – 70% or 78% – 71% at a range of $0 – $200 000 WTP using PSA thresholds≥ 2 ng/mL and ≥ 4 ng/mL, respectively
Govers et al [145]	SelectMDx v PSA	-$1,694 (-$1,854)	0.045	Dominates	NR
Dijkstra et al [147]	SelectMDx v PSA	-€128 (-$170)	0.025	Dominates	NR
Schiffer et al [141]	UPA-PC v PSA	-€297 (-$440)	NA as study was cost-consequence analysis	NA	NR
Kim et al [142]	MRI + biopsy only if PHI ≥ 30 v MRI + biopsy for all	-£191 (-$280)	NA as study was cost-consequence analysis	NR	NR
Teoh et al [144]	PHI v PSA	-$4562 (-$4657)	0.35	Dominates	NR
Karlsson et al [143]	Stockholm3 if PSA > 2 ng/ml v PSA	€14 ($18)	1	€5663 ($7082)	97% at WTP €50,000
Govers et al [139]	SelectMDx v PSA	France: -€1217 (-$1620) Germany: -€439 (-$605) Italy: -€757 (-$1089) Spain: -€247 (-$405)	France 0.036 Germany 0.026 Italy 0.043 Spain 0.028	Dominates	NR
Venderink et al [154]	MRI TRUS fusion biopsy v TRUS biopsy	€175 ($236)	0.1263	€1386 ($1869)	NR

Cerantola et al [155]	MRI cognitive-targeted biopsy v TRUS biopsy	-CAD$2,187 (-$1,960)	0.168	Dominates	NR
de Rooij et al [156]	MRI targeted biopsy v TRUS biopsy	€31 ($42)	0.10	€323 ($442)	80% at WTP higher than €2000
Faria et al [152]	mpMRI guided biopsy v TRUS biopsy	NR	NR	£7,076 ($10,519)	NR
Pahwa et al [153]	MRI cognitive-targeted biopsy v TRUS biopsy	-$1771 (-$1882)	0.198	Dominates	94.05% at WTP $50,000 and 93.9% at WTP $100,000
Mowatt et al [159]	T2-MRI vs TRUS biopsy	£7 ($12)	0.00054	£12,315 ($21,013)	34% at WTP £30,000
Barnett et al [151]	Combined (standard + targeted fusion) biopsy v TRUS biopsy	NR	3.5	$23,483 ($24,340)	NR
Nicholson et al [158]	clinical assessment + mpMRI v clinical assessment	£113,449 ($180, 497)	3.35	£33,911 ($53,952)	100% at WTP £37,000
Barnett et al [150]	hybrid 18F-choline PET/mpMRI with Likert scoring v TRUS biopsy	NR	1.1	$35,108 ($35,841)	NR
Callender et al [28]	MRI-first risk-stratified screening at 10-year absolute risk threshold of 7.5% v no screening	£28 ($35)	0.0042	NR	NR

**NA indicates not applicable as study was not a cost-utility analysis, NR indicates not reported as study did not report differences between interventions.*

[¥]Costs are in reported currency with USD 2020 costs in brackets to aid comparison

[∞] Where more than two interventions were compared the ICER for the most cost-effective intervention is presented.

Assessing uncertainty in cost-effectiveness results

Five studies found that the results were sensitive to the potential of the tests to identify cancer, particularly clinically significant cancer.[28 145 147 152 153 156] For example, Govers et al found that if the sensitivity of the SelectMDx strategy for detecting clinically significant disease decreased from 95.7% to 80.4% the strategy would no longer improve health outcomes in terms of QALYs gained. Three studies found that the results were sensitive to the assumed prevalence of cancer and significant cancer.[153 154 156] Pahwa et al found that using a higher Gleason score cut off, implying a higher threshold for a cancer to be deemed clinically significant, increased cost-effectiveness across all strategies. The cost of the tests

was also stated as an important factor in five of the studies [140 143 146 152 154], although in these cases the estimated costs would have to change substantially to impact the results. Bouttell et al, for example, found that the cost of the PHI test would have to be HK$8,500 (base case HK$3,000) for the PHI strategy to be cost neutral rather than cost saving and Venderink found that the ICER of MRI-TRUS fusion would be higher than the WTP threshold if mpMRI cost €9500 ($10,000) or more per patient instead of €317.

Two studies found that the results were sensitive to the probabilities of cancer progression in undiagnosed cases.[151 157] Barnett et al stated that the cost per QALY gained relative to no screening only exceeded the WTP threshold when it was assumed that the risk of developing metastases in undiagnosed cases was much lower than that in the base case. Two studies found the survival rate to have an influence on results.[154 157] The NICE guideline stated that increasing the survival rate resulted in the strategy where all men receive an immediate TRUS and no subsequent follow-up to be optimal in the majority of sub-populations. In contrast, Venderink et al found that if the yearly survival rate among patients with treated clinically significant prostate cancer were to decrease from 98.6% to 93.2%, or from 99.2% to 96.5% for untreated insignificant prostate cancer, TRUS biopsy would be the most cost-effective strategy.

Finally, two studies found the utility levels for diagnosed cancer states to be influential.[151 159] Mowatt et al, for example, assessed the impact of applying a utility decrement of 0.035 (half of the disutility associated with having moderate anxiety rather than no anxiety on the EQ-5D) to patients with undiagnosed cancer, to reflect potential disutility from increased anxiety associated with having a high PSA but no diagnosis. The sensitivity analysis to utility values resulted in systematic TRUS being the most cost-effective intervention.

3.4. Discussion

The aim of the review was to identify model based economic evaluations evaluating new diagnostic tests for prostate cancer; determine the evidence base and cost-effectiveness

results; provide an overview of the characteristics of these models and their data sources to aid in the cost-effectiveness analysis carried out as part of this dissertation; and assess the limitations of available models, providing guidance on future improvements. Twenty-two studies were identified, all published between 2011 and 2021. Eleven compared the cost-effectiveness of new urinary or blood biomarkers to each other or to the standard of care (a PSA test). Another eight compared different approaches to prostate biopsy, and three compared follow-up strategies in men who have a negative initial biopsy result. Most models used either a combined decision tree/Markov or purely Markov model structure with only seven modelling progression through stages or grades of cancer. Substantial variability was seen in the model pathways of prostate cancer natural history; the data sources used to inform progression; treatment allocation assumed for high and low risk cancers; disutility values assigned to health states; and the assumed accuracy of the tests. All but one study [158] found the introduction of these novel tests to be cost-effective; however, in some cases the benefits may be overestimated due to a failure to take account of overdiagnosis and the natural history of the disease in untested men.

3.4.1. Recommendations for a future cost-effectiveness model

This review has highlighted that many models fail to account for the complexity of the disease including stage or grade of cancer and how cancer progresses in diagnosed and undiagnosed cases. It has shown that a new model should consider the entire diagnostic pathway from testing to treatment to comprehensively assess the 'true' cost-effectiveness of these tests within a diagnostic strategy for prostate cancer. When modelling the lifetime cost-effectiveness of a test to diagnose prostate cancer, it is important to consider the natural history of the disease, and how a test may impact on this, to ensure that the benefit of the test is accurately represented and overdiagnosis is considered. The studies identified in this review which modelled the natural history of prostate cancer all did so in different ways, suggesting a lack of clarity in the field. Any future model should consider long-term outcomes, ensuring the natural history of the disease is accurately modelled, and taking progression through stages and grade of prostate cancer into account.

Although a formal literature review to identify health state utility values has not been carried out, the values used in previous models indicate a potential paucity of information on how prostate cancer treatment and adverse effects impact on quality of life. If up-to-date EQ-5D surveys of men with prostate cancer are unavailable to future analysts, uncertainty in these estimates should be fully accounted for using probabilistic and deterministic sensitivity analyses, as this could greatly impact the results of a cost-utility model. Further, the review has also highlighted the failure of many models to fully account for uncertainty in other model parameters. Future models should ensure that uncertainty is represented by assigning uncertainty distributions to all uncertain parameters. The parameters having the most significant impact on cost-effectiveness results should also be considered using Value of Information techniques.[230]

3.4.2. Strengths and Limitations of Review

The strength of this systematic review is that it has provided an overview of cost-effectiveness analyses published in the last ten years which have compared novel diagnostic methods in prostate cancer. It has offered insight into the data parameters that will be needed to populate a future cost-effectiveness model incorporating new tests and diagnostic strategies in prostate cancer, and potential sources of information for these parameters. It has also highlighted the limitations of previous models. The results from the review have emphasized the importance of accurately estimating factors such as the sensitivity of tests, the prevalence of disease and the progression of the disease.

A limitation is that this review cannot provide recommendations on the *most* cost-effective test or diagnostic strategy as the studies are too heterogeneous in setting and strategies compared. A further limitation is that, although the systematic review did not identify any relevant studies published between 2009 and 2011, the 2009 cut-off could potentially miss cost-effectiveness models of novel diagnostic methods published prior to 2009. A further limitation is that, in the rapidly evolving context of prostate cancer screening, the review may already be out of date having been carried out in 2021.

3.4.3. Comparison with previous reviews

No previous systematic review could be identified considering cost-effectiveness models focused on screening and diagnostic strategies beyond standard PSA-based testing. One recent systematic review assessed model-based economic evaluations of PSA-based screening strategies only.[202] This review also found significant variation in model pathways to reflect cancer progression in the ten included studies and limited and heterogenous evidence on quality of life. Three older reviews were also identified but all assessed PSA-based screening only.[231-233]

3.5. Conclusions

The introduction of new biomarkers and biopsy methods involving MRI in the studies identified in this review has been shown to lead to an improvement in health outcomes and a decrease or acceptable increase in costs.[140 145 147] Current concerns around implementing PSA-based prostate cancer screening strategies are due to overdiagnosis and overtreatment [234], and these newer methods may lead to a reduction in these factors. This review has highlighted the substantial complexity involved in modelling the cost-effectiveness of diagnostic tests in prostate cancer to determine whether these strategies should be used at all and if so, how and in what combination. To ensure the cost-effectiveness of any diagnostic strategy is assessed robustly, there is a need to consider long-term outcomes, ensuring disease progression in diagnosed and undiagnosed cases is accurately represented, uncertainty is fully accounted for, quality of life estimates are measured as accurately as possible, and the possibility of repeat screening and testing in men with a negative diagnosis is considered.

CHAPTER 4. GAINING CONSENSUS ON UK RELEVANT PROSTATE CANCER SCREENING STRATEGIES

4.1. Introduction/Background

Economic decision analytic modelling can compare the costs and consequences of a screening strategy and the resulting care pathway over a person's lifetime. As mentioned in section 1.1, the first step in any economic evaluation, however, is to determine the decision question, which includes identifying all relevant comparators or screening strategies. As the frequency of screening, population to screen, diagnostic test or tests used, and their order and combination are all subject to changing practice, this can make identifying all possible comparators, especially what is relevant, challenging.

As mentioned in section 1.1 and 2.2.10, prostate cancer screening in particular is an area where considerable uncertainty as to what the comparators in a cost-effectiveness analysis should be exists. Recent developments such, as the possibility of risk stratification,[235] potentially better early detection biomarkers e.g. STHLM3,[236] and new diagnostic strategies (including the use of multiparametric MRI[25]), offer opportunities for improving the outcomes of prostate cancer screening. However, it is not yet clear how or if these developments should be implemented. In addition, there is uncertainty around other factors, such as at what age and how often men should be screened, as shown in section 2.2.7. There is therefore an ongoing debate around whether to screen and if so, who, how and how often in an era of rapid evolving practice and advancements in screening and testing technologies.

4.1.1. Current recommendations on how to choose comparators in a cost-effectiveness analysis

Generally, economic evaluations are recommended to include all possible comparators, or clearly justify their exclusion.[12 237 238] This is important, as the selection of comparators may change the conclusions of the analysis.[239] Most guidelines recommend that authors consult with experts to ensure their choices are relevant. For example, Philips et al in their good practice guideline on Decision-Analytic Modelling in Health Technology Assessment recommend that comparators are based on 'literature reviews and expert opinion'.[12] Similarly, the report by the ISPOR-SMDM Modelling Good Research Practices Task Force on conceptualizing a model states that 'the modelling team should consult widely with subject experts and stakeholders to ensure that the model represents disease processes appropriately and adequately addresses the decision problem'.[240] In addition, a report by the Health Economics and Decision Science group[241] on conceptual modelling for health economic model development states 'the question of "what is relevant?" to a particular decision problem should not be judged solely by the individual developing the model; rather making such decisions should be considered as a joint task between modellers, decision-makers, health professionals and other stakeholders who impact upon or are impacted upon by the decision problem under consideration.' Failing to account for a wide range of views may lead to the development of a model and production of results which are not useful to the decision makers or people currently working in the field. Yang et al state that comparators can be justified by carrying out 'a horizon scan of the diagnostic landscape with clinical experts' opinions'[238], although what would be considered a 'horizon scan' or 'expert' opinion is not defined.

4.1.2. Aim

The aim of this chapter was to address decision question uncertainty in prostate cancer screening by eliciting the most relevant strategies to compare in a cost-effectiveness model from a range of experts working in the field. The goal was to ensure the applicability and importance of the eventual cost-effectiveness results. This work has been published as a manuscript in Pharmacoeconomics.[242]

4.2. Methods

4.2.1. Modified-Delphi technique

A modified-Delphi technique was chosen to gather consensus views. This is a consensus-based technique that provides a systematic method of collecting and aggregating informed judgments from a group of experts via multiple iterations.[243] It requires expert participants to provide their opinions in sequential questionnaires (rounds), with each round presenting group feedback from the previous round. Feedback from sequential rounds encourages participants to reassess, alter and/or develop opinions. Anonymity of the responses is maintained to ensure that no individual dominates the process.[244]

A conventional Delphi process begins by consulting an expert panel on a broad question to generate answers that will then be rated or ranked in the subsequent round of questioning. Using a modified Delphi designs means that, rather than consulting an expert panel to generate answers to the round 1 question(s), the researcher collects the initial answers to the question(s) through some other means, such as a literature review, and presents them to the panel to begin the consensus seeking process.[245] Although the Delphi method was originally intended to obtain the most reliable consensus of opinions of a group of experts, Landeta 2006 state that 'later applications of the Delphi method have eliminated the restriction of the obligatory search for consensus, so that today it might be defined as a social research technique whose aim is to obtain a reliable group opinion using a group of experts.'[244]

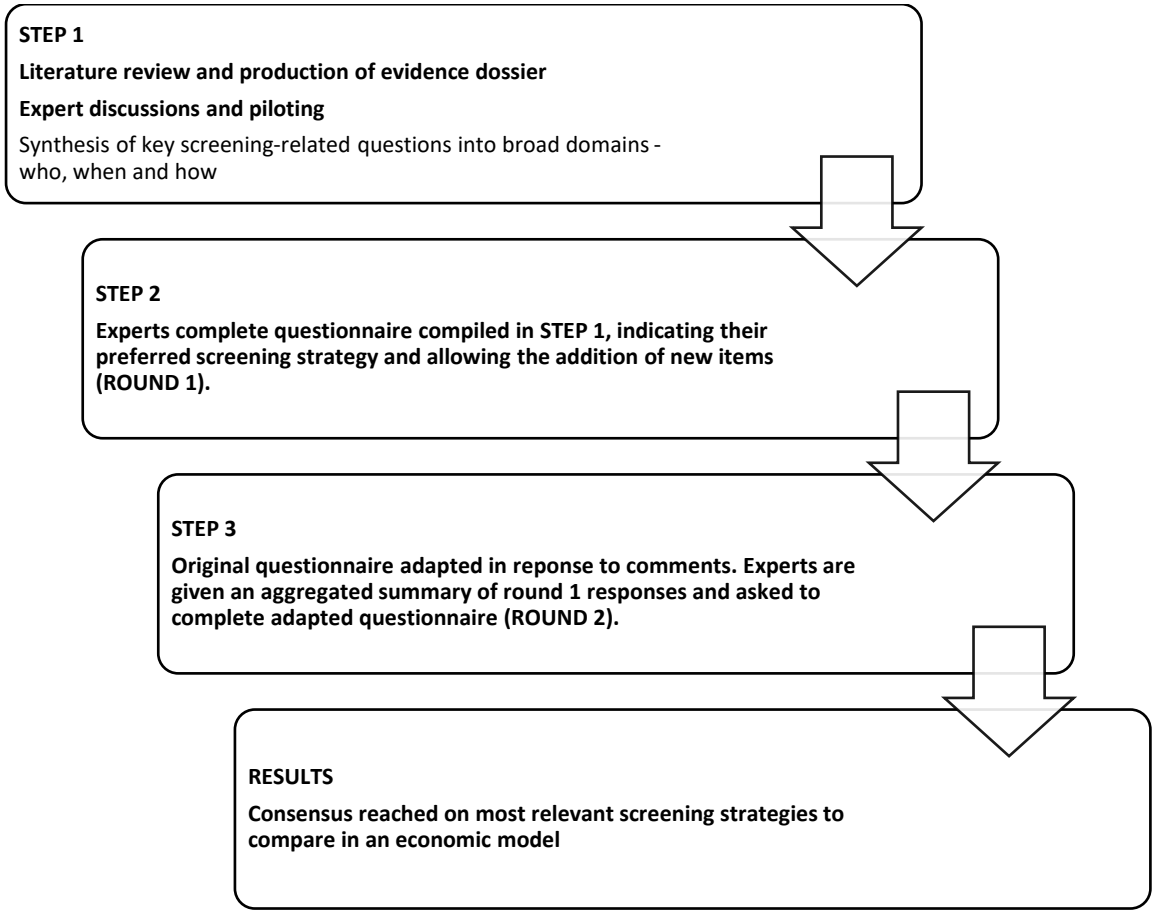
The Delphi method has been extensively and effectively used to obtain consensus within the realm of prostate cancer screening and treatment across countries and participant backgrounds.[246-248] Aims of the Delphi process in this context have included to form recommendations on how and when to use the PSA test[246], how genetic counselling and testing should be incorporated in prostate cancer screening and management[247], and how multiparametric MRI should be used in the interrogation of prostate neoplasia in clinical practice and focal therapy.[248] Various techniques have been used to recruit participants

including using UK databases to contact a wide selection of hospital doctors and nurses working within urology or oncology, GP practice managers and Clinical Commisioning Group Clinical Leads,[246] purposively selecting an expert panel based on achieving a balance of expertise,[247] and using a systematic literature search to identify experts based on authorship or peer-recommendation.[248] The process has varied from three online questionnaire rounds and an in-person meeting[246 248] to one round of in-person anonymous voting followed by re-administering select questions where there was debate among panellists.[247] Consensus has been defined as an agreement of at least 70% from the respondents[246 248] or ≥ 75% for strong consensus and 50% to 74% for moderate consensus.[247]

Other methods of collecting opinion exist such as interviews, focus groups and the nominal group technique. The other methods involve face-to-face discussions, with a moderator appointed to ensure that all individuals can contribute.[249] The benefit of in-person discussions is that they allow the decisions and reasoning behind them to be recorded in real time.[240 250] Due to the simple format of focus groups, it is possible that they could be used to facilitate the involvement of patient representatives, in addition to clinicians and economists. However, a potential drawback is the practicality of organising and running face to-face discussions with stakeholders.[251] The Delphi method was chosen in this case as it maximises the benefits of using an expert panel while minimising potential disadvantages by implementing anonymity.[243] The method also allowed the study to be conducted in geographically dispersed locations without physically bringing the respondents together, saving time and money. Another advantage was that participants had time to think through their ideas without the pressure of an in-person group meeting and could respond in their own time. The results were also easier to review and analyse than those from a meeting or focus group discussion.[252]

The modified technique used in this case is described in Figure 3.

Figure 3. Modified Delphi technique



4.2.2. Participants and Recruitment

In identifying experts, it was first decided what type of expertise was needed to answer the questions (i.e. expertise in prostate cancer early detection, treatment, modelling or the role of genetics in prostate cancer risk), and then individuals with this particular expertise were identified. Participants were identified using purposive sampling, which focuses on the views of those able to provide in-depth knowledge of the topic of interest.[253] In-depth knowledge related to factors such as on-the-job experience and peer-reviewed academic output. The participants worked in a variety of institutions including Universities in the UK, Sweden, Finland, the Netherlands and the US, cancer centres in the US, NHS trusts, Public Health England, and the Institute of Cancer Research. The selection of participants from a variety of backgrounds and institutions was important to encourage diversity of opinions.

Snowball sampling was also used, whereby the contacts were asked to suggest others within or outside their organisations that might offer insight.[254] To facilitate follow-up and establish a response rate, the participants were asked to provide a list of email addresses to which the questionnaire might be sent rather than forwarding the questionnaire themselves. Prior research experience, clinical focus, country, and age range were included in the questionnaire so that comparative analysis could be carried out.

There is no statistical basis on which to determine the necessary sample size for a Delphi survey. Previous studies have shown that reliable outcomes can be obtained with a Delphi panel consisting of a relatively small number of Delphi experts (n=23) selected via strict inclusion criteria.[255] The questionnaire was therefore sent to 20 experts in the first instance, with 7 additional participants included through snowball sampling. Approval for the study was granted by the Faculty of Health Sciences Research Ethics Committee of the University of Bristol (Approval reference number: 91622).

4.2.3. Step 1. Production of questionnaire and evidence dossier

To identify the extent of the uncertainty in the decision question, a rapid review of the literature was carried out. This review involved identifying and summarising current UK and international guidelines on prostate cancer screening and recent large trials that have informed these guidelines and which addressed different aspects of screening. This review, which was intended to provide a rapid overview rather than a detailed description of all relevant trials, was summarised in an evidence dossier. The evidence dossier was circulated to an advisory panel comprising two oncologists, two urologists, one GP, and two researchers in prostate cancer screening and clinical oncology (four from the UK and three from the US), to identify any missing evidence. As a systematic search of the literature was not carried out, the expert panel was asked to review the evidence dossier and highlight any missing information that may be relevant. The selection of a group from varied clinical and geographical backgrounds avoided any key pieces of information being overlooked.

The findings informed the development of the questionnaire, which included questions centred around three primary domains of uncertainty identified in the evidence dossier: the group of men that should be invited for screening, how often they should be screened, and which diagnostic procedures to use. Although training was not provided to the participants in advance, the questionnaire was piloted for language, comprehension and ease of use on a smaller group comprising a clinical oncology researcher, a medical oncologist, a GP and a urologist. No difficulty was demonstrated in completing the questionnaire and the answers given indicated a good degree of understanding. Changes were made to the questionnaire in response to their comments, such as a move away from a ranking of preferences within each question and towards the selection of one preferred approach. This enabled the experts' preferences to be more clearly drawn out and was in response to comments that after a certain rank the experts had no real preference.

4.2.4. Step 2: Dissemination of questionnaire

A Web administered survey was developed using REDCap electronic data capture tools hosted at the University of Bristol.[256] Links to the questionnaires and evidence dossier were circulated via email. After reading the participant information sheet detailing the aim of the study, why individuals were being asked to take part, what the process would involve, and what would happen with the results, participants were asked to provide their informed (online) consent to proceed with questionnaire completion. The participant information sheet is shown in Appendix 6.

The first round was completed November 2019, and the second round January 2020. Approximately 2 weeks after distributing both questionnaires reminder emails were sent to try and increase response rates. In round 1, respondents were asked to indicate their preferred screening strategy through a series of questions. Completion of these questions resulted in an automatic statement being generated for each respondent that summarised their answers. An example is shown below:

> *"Your preferred screening strategy is age-based, in men aged 50 to 75. You believe the PSA threshold for further investigation should be 3 ng/ml and that the screening interval should be every two years. Men with a raised PSA should be tested using a multi-kallikrein panel (e.g. 4k score, STHLM3) and multiparametric MRI (mpMRI) prior to biopsy."*

Using a free-text section at the end of each question, respondents were able to add items they considered important but that were not already covered. An example of a completed questionnaire is available in Appendix 7.

4.2.5. Step 3: Updating of questionnaire

The first-round responses informed the content and modification of the updated questionnaire in order for participants to then repeat the questionnaire a second time (round 2). At the end of round 1, the percentage of participants choosing each item was summarised. Items were not considered in round 2 if not chosen by any participant in round 1. New items were considered if suggested by more than 10% of participants. This is in line with a previous Delphi consensus process carried out to rank core items of resource use that should be included in economic evaluations of health care interventions.[257] This updated questionnaire, in the form of 13 consensus statements, was then sent to the same key contacts from round 1 and their lists of suggested contacts. Feedback from round 1 was presented for each round 2 item in the form of the number of participants choosing each option. As 2 months had passed since the first round, participants were given a reminder of their own choice. Comments from round 1 were also summarized and presented.

As is common in Delphi procedures[258 259], respondents were requested to rank their agreement with the 13 statements on a 9-point Likert scale. Consensus on a statement was considered reached if scored 7 to 9 (moderately agree, agree, or strongly agree) by more than 70% of participants and 1 to 3 (strongly disagree, disagree or moderately disagree) by less than 15% of all participants. In place of conducting further Delphi rounds, which were not deemed necessary after considering the results of the first two rounds, more stringent criteria were also set (>70% scoring an item 8 or 9 and <15% scoring 1–3) to aid discussion on the most important aspects to participants. This was in line with the approach taken in a previous consensus process carried out by Thorn et al.[257] Within-group interrater agreement was assessed using the *r*wg* statistic with the rectangular null and maximum dissensus null distributions.[260] This is a commonly used statistic to quantify consensus in ratings on a Likert

scale.[261] With this statistic, r*wg ≥ 0.80 may be considered high enough agreement to establish interrater agreement with 10 or more raters.

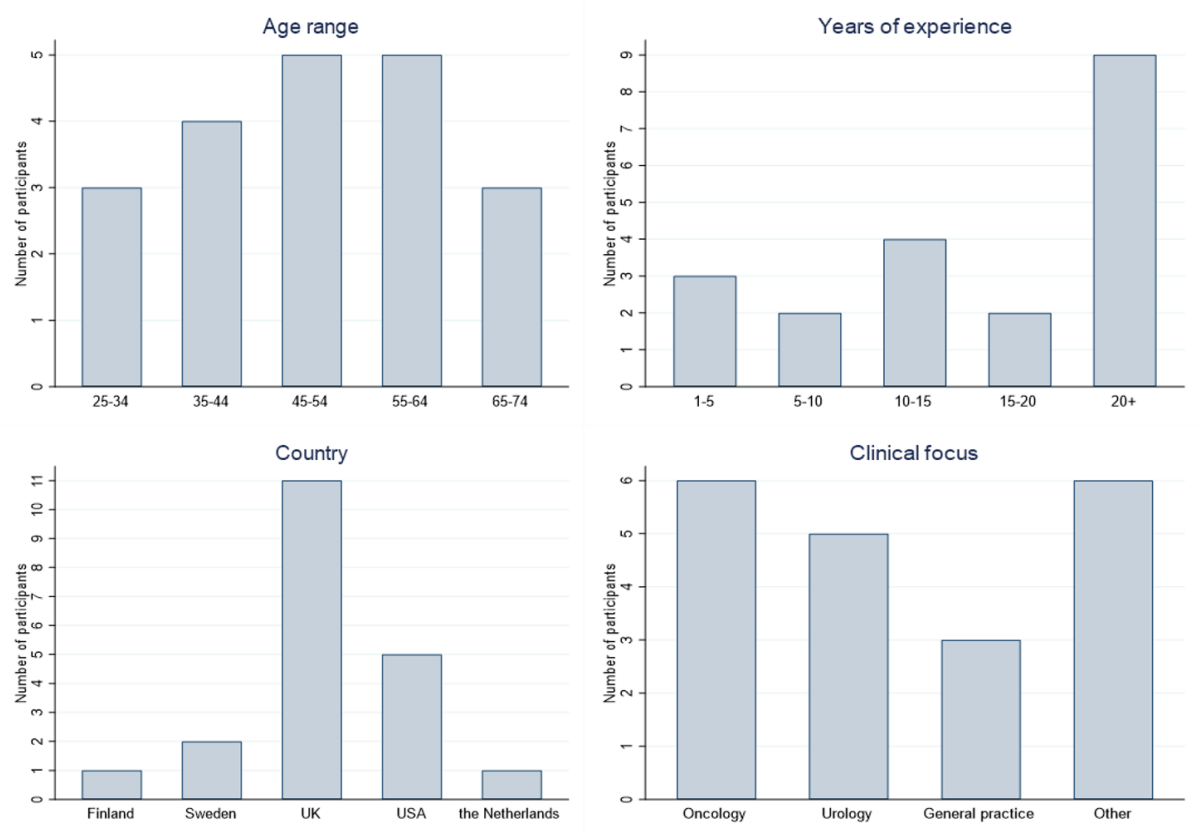
4.3. Results

4.3.1. Round 1

Characteristics of respondents

Twenty participants responded to the questionnaire out of 27 invitees, giving a response rate of 74%. The respondents were of varying ages with half being in the 45-64 age bracket. Nine out of 20 (45%) respondents had over 20 years' experience in their field. Responses were received from 6 oncologists, 5 urologists, 3 GPs, 2 public health specialists, 2 researchers and 2 people involved in national screening services. Four of the respondents had extensive experience in cancer epidemiology, and three of these were Professors in general epidemiology or in cancer epidemiology, specifically. Just over half of the respondents were from the UK (11/20, 55%) (Figure 4). As responses were received from participants with a range of expertise and backgrounds, 20 participants was considered to be enough to collect rich data and allow exploration of the screening strategies thought to be important.

Figure 4. Characteristics of respondents



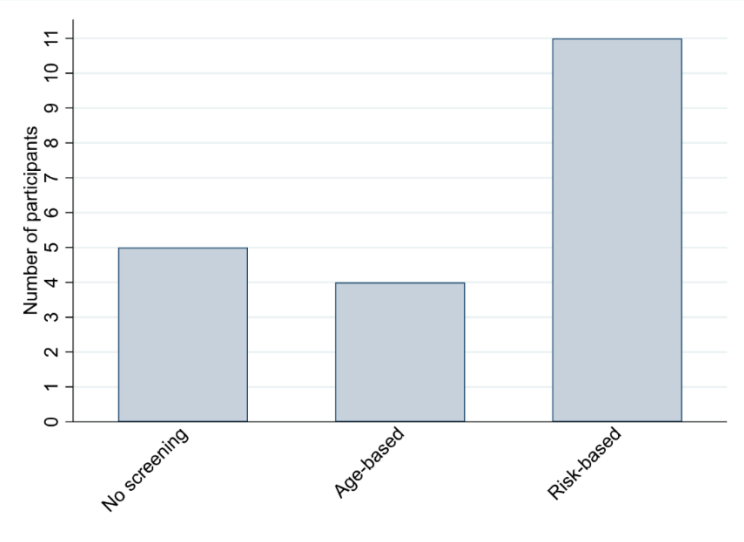
Domain 1: Who should be invited for screening?

The first question asked what type of prostate cancer screening programme participants felt should be provided. Options included no screening, opportunistic screening, organised age-based screening and organised risk-based screening. No participant chose opportunistic screening. The most popular choice was risk-based screening, chosen by 11 participants (55%), followed by no screening (5 participants/25%) and age-based screening (4 participants/20%) (Figure 5).

Risk-based screening was chosen by all five urologists and no screening was chosen by all of those involved in national screening services. No other trends were identified based on age, experience, or country of residence. Reasons given for preferring risk-based screening included targeting men who are more likely to benefit, thus avoiding the harms of age-based screening, “early detection and intervention for people with strong risk factors”, “more scope for improvement as we learn to better estimate a man's risk”, and “avoiding the unnecessary costs and sequelae of universal screening”. All participants who selected

'no screening' highlighted that this was due to a lack of current evidence to suggest that screening offers more good than harm. Reasons given for selecting age-based screening included preventing opportunistic screening in the “wrong” age ranges, and for simplicity.

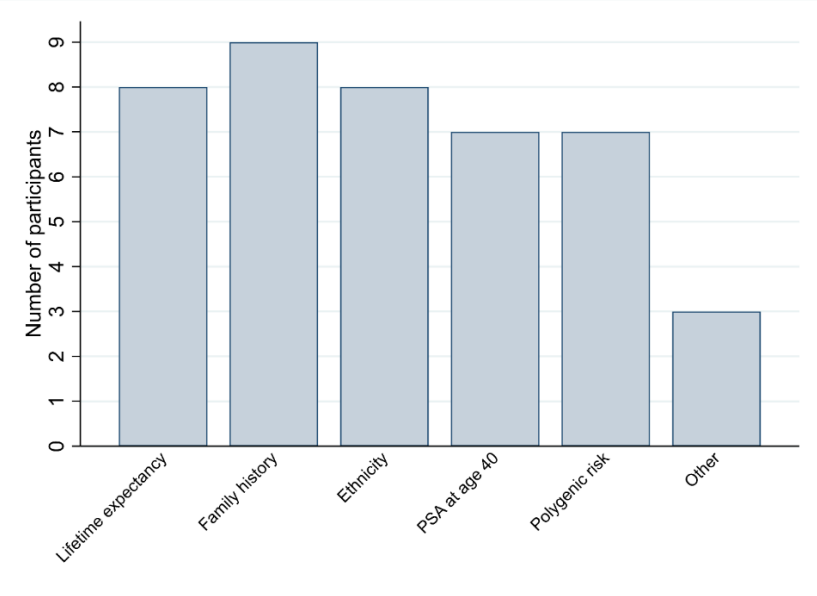
Figure 5. Number of participants selecting each response to question “What type of prostate cancer screening programme do you feel should be provided in the UK?”



To get agreement on what the participants considered to be important risk factors, the second question asked all participants who had indicated risk-based screening to be their preferred option which factor/s they thought risk should be based on. The options included lifetime expectancy >10 years, family history, ethnicity, PSA at age 40 above a certain threshold, and polygenic risk. Multiple options could be chosen. 5 of 11 participants who chose risk-based screening ticked all available options. The most popular option was family history (chosen by 9 participants), followed by lifetime expectancy (chosen by 8) and ethnicity (8) (Figure 6). PSA at age 40 and polygenic risk were also chosen by 7 participants, giving an even spread of responses. Suggestions given in the 'other' column were biomarkers and germline high risk mutations.

Comments on this question centred around the understanding that "all available risk factors which can be measured reliably and affordably should be included" with the optimal combination of factors being unclear.

Figure 6 Number of participants selecting each response to question "If risk-based screening were to be provided what factor/s should risk be based on?"

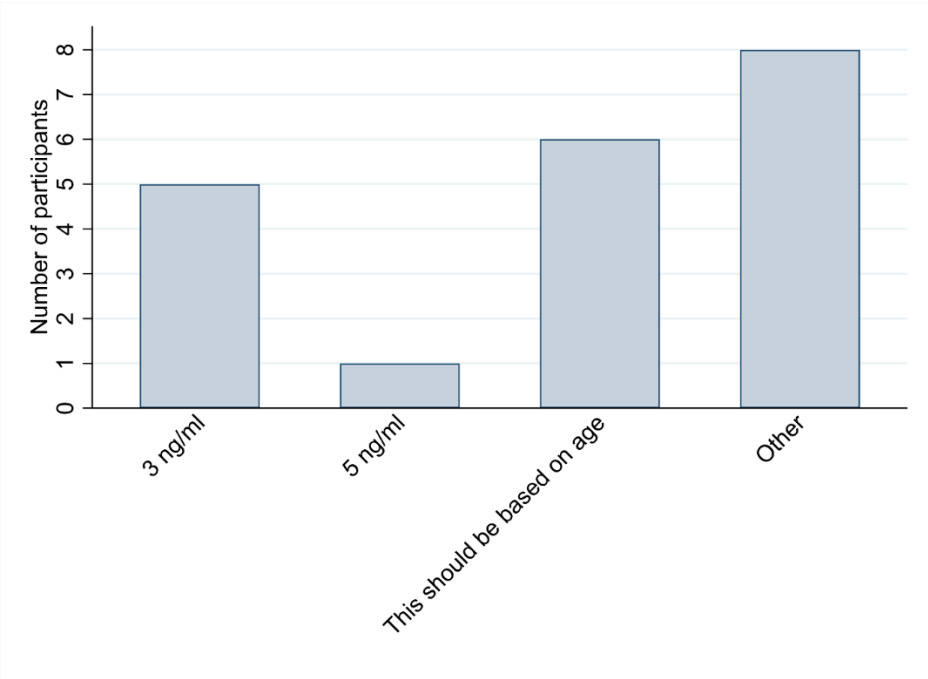


Domain 2: Which diagnostic procedures?

The next question asked participants which PSA threshold they thought should be used to indicate further investigation with the options being: 3 ng/ml, 4 ng/ml, 5 ng/ml, 6 ng/ml, 7 ng/ml, 8.5 ng/ml, 10 ng/ml, and 'this should be based on age'. No participants chose 4 ng/ml, 6 ng/ml, 7 ng/ml, 8.5 ng/ml or 10 ng/ml. 5 ng/ml was chosen by 1 of 20 participants (Figure 7). In the 'other' column, 2 participants suggested that the PSA threshold for investigation should be lower than 3 ng/ml, 2 suggested that PSA should be combined with other biomarkers, 2 suggested that risk calculators should be used in place of PSA, and 2 commented that they did not support screening.

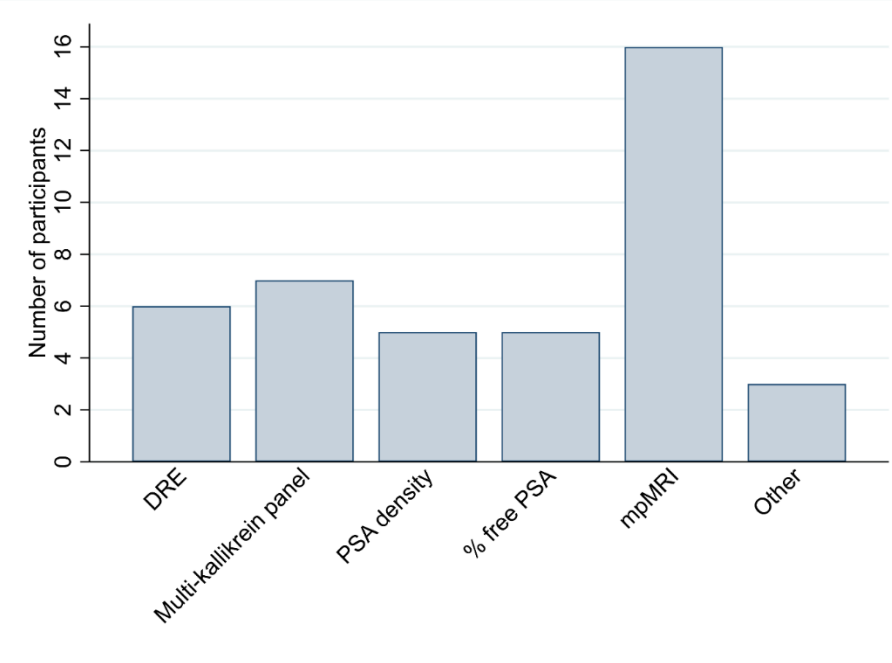
Comments on this question generally reflected the idea that "a sequence of tests" should be carried out before proceeding with a biopsy. One participant commented, "An initial test solely based on PSA is good idea, since a low PSA value has a good negative predictive value. If a man has PSA higher than, say, 1.5 or 2 further testing can be done using a more refined test, e.g. Stockholm3, PHI or 4K." Although 5 people chose a level of 3 ng/ml as their preferred option, 3 people commented that trials have shown using a level of 3 ng/ml will mean that some aggressive cancers are missed.

Figure 7 Number of participants selecting each response to question "Which PSA threshold do you think should be used to indicate further investigation?"



The next question asked respondents what further investigation(s) they thought men with a raised PSA level should have prior to being offered biopsy. The options included were no further investigation, digital rectal examination, a multi-kallikrein panel (e.g. 4kscore, STHLM3), PSA density, % free PSA, and multiparametric MRI (mpMRI). Multiple options could be chosen. The most popular option was mpMRI with 16 respondents (80%) indicating that they thought this should be used prior to biopsy. 7 out of 20 (35%) respondents indicated that a multi-kallikrein panel should be used, either with mpMRI alone (4 participants) or alongside PSA density, % free PSA and mpMRI (3 participants). 6 chose digital rectal examination as an option (Figure 8). Of those who chose 'other' the responses were 'no screening' and 'polygenic markers'. Several people commented that the optimal approach would be the "use a good reflex test and then MRI as the next step", with a multi-kallikrein panel and/or PSA density or % free PSA being suggested as reflex tests.

Figure 8. Number of participants selecting each response to question "What further investigation should men with a raised PSA level have prior to being offered a biopsy?"

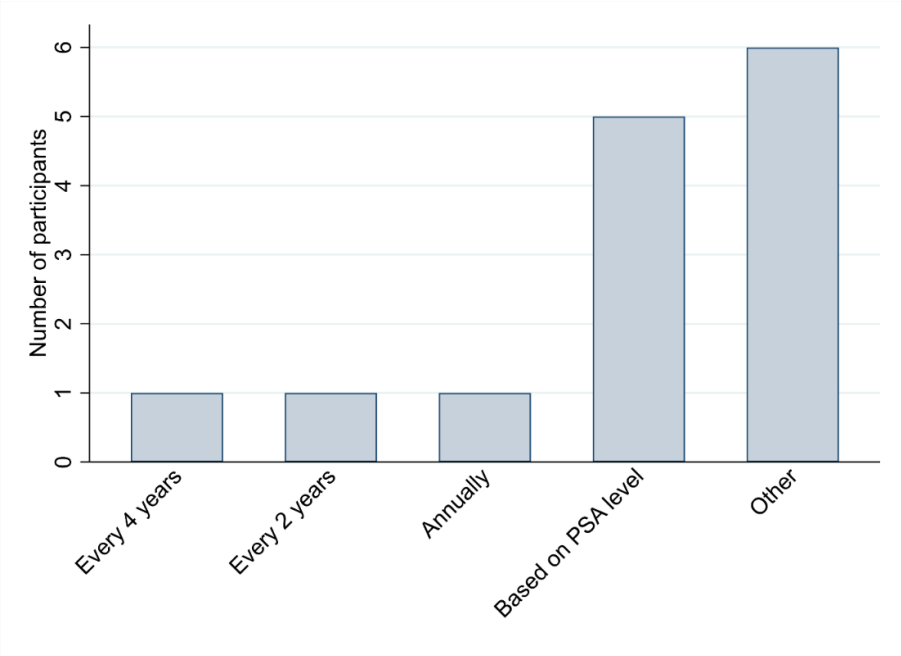


Domain 3: How frequently to screen?

The final question asked participants, given their optimal screening strategy, how often they thought men should be screened. The options given were every 10, 6, 4, and 2 years, annually, only once, and 'this should be based on PSA level'. No participants chose 10 or 6 years as an appropriate screening interval and only 1 participant each chose 4 years, 2 years and annually. The most popular responses were 'this should be based on PSA level' and 'other'.

Of those who thought that screening interval should be based on PSA level, all stated that men with a PSA level less than 3 ng/ml could be screened every 2-3 years whereas men with a higher PSA should be screened annually. In the 'other' column the suggestions made were mainly around screening interval being based on risk. Comments included "Testing frequency should be based on risk. For low-risk men, screening every 4-6 years is likely enough. Some high-risk men (a small fraction) should probably return every year for screening", and "a risk model that includes PSA, age and other risk factors (e.g. ethnicity) might be more sensitive and provide a better follow up schedule".

Figure 9. Number of participants selecting each response to question "Assuming some optimal strategy for inviting men to be screened has been adopted, how frequently do you think men should be screened?"



In summary, the responses to round 1 indicated several items that were not of interest to compare in a cost-effectiveness model e.g. opportunistic screening, a PSA threshold for further investigation higher than 3 or a fixed screening interval for all men, as these were not chosen by any participant. On the other hand, a clear consensus (in terms of the pre-defned criteria of >70% agreement) was shown towards the use of mpMRI prior to biopsy.

4.3.2. Round 2

The results from the first round led to the generation of a list of 13 statements to help clarify the respondents' preferences. The statements that participants were asked to indicate their agreement with and the results are shown in Table 8. Items which were chosen by less than 10% of respondents in round 1 were not questioned further nor were items which were chosen by more than 70% of respondents (namely that mpMRI should be used in men with a raised PSA level prior to biopsy), as consensus was already considered reached. Seventeen of twenty participants responded to round 2 giving an 85% response rate. The participants were again from a wide and representative range of backgrounds. As interest was shown in all three of no screening, age-based and risk-based strategies, respondents were asked their opinion on each but this time in the context of their inclusion as comparators in a cost-effectiveness model. Similarly, further statements attempted to clarify the respondent's views on the inclusion of PSA testing in the screening pathways to be considered in a cost-effectiveness model and how screening intervals should be determined.

This table shows the percentage of participants rating a statement 7-9 (7 being moderately agree, 8 being agree, and 9 being strongly agree). A threshold of 70% in this category was the pre-agreed marker of consensus. The percentage of patients rating a statement 8-9 is also shown, as well as the percentage rating a statement 1 to 3 (strongly disagree, disagree or moderately disagree).

Table 8. Final outcomes for statements after round 2

Statement	% rating 7-9	% rating 8 or 9	% rating 1-3	Outcome: pre-agreed rules	Outcome: more stringent rules	Interrater agreement (r^*_{wg})
1. It is useful to compare the cost-effectiveness of no screening to other screening strategies in a future economic model	88.24%	88.24%	0.00%	Consensus reached	Consensus reached	0.90
2. It is useful to compare the cost-effectiveness of inviting all men within a certain age range to be screened to other screening strategies in a future economic model	94.12%	82.35%	5.88%	Consensus reached	Consensus reached	0.89
3. If it is possible to identify men at higher risk of developing prostate cancer prior to testing (through the use of polygenic risk scores, family history, ethnicity or otherwise), it would be useful to compare the cost-effectiveness of inviting only higher risk men for screening	88.24%	82.35%	0.00%	Consensus reached	Consensus reached	0.94

4. If it is possible to identify men at higher risk of developing prostate cancer prior to testing, it would be useful to compare the cost-effectiveness of inviting all men within a certain age bracket for screening but screening higher risk men at an earlier age	88.24%	70.59%	5.88%	Consensus reached	Consensus reached	0.82
5. PSA in isolation should no longer be used as a reflex test to trigger MRI/prostate biopsy	35.29%	23.53%	17.65%	Consensus not reached	Consensus not reached	0.75
6. A PSA test should be used before a more sophisticated biomarker or risk model (e.g. 4k score, STHLM3) and only men with total PSA above a certain threshold should be tested using the biomarker or risk model	41.18%	35.29%	23.53%	Consensus not reached	Consensus not reached	0.65
7. A PSA threshold of 1.5 ng/ml has enough negative predictive value to exclude any further testing	35.29%	23.53%	17.65%	Consensus not reached	Consensus not reached	0.65
8. The threshold for further investigation should increase as men age	47.06%	35.29%	5.88%	Consensus not reached	Consensus not reached	0.76

9. It would be useful to assess the cost-effectiveness of using a multi-kallikrein panel or risk model (e.g. 4k score, STHLM3) as a reflex test to triage patients suitable for mpMRI prior to biopsy	64.71%	52.94%	0.00%	Consensus not reached	Consensus not reached	0.87
10. It would be useful to assess the cost-effectiveness of using PSA density and % free PSA alongside a multi-kallikrein panel as reflex tests to triage patients suitable for mpMRI prior to biopsy	52.94%	41.18%	0.00%	Consensus not reached	Consensus not reached	0.89
11. All men being screened should be offered a DRE	17.65%	17.65%	47.06%	Consensus not reached	Consensus not reached	0.59
12. If it is possible to identify men at higher risk of developing prostate cancer (through the use of polygenic risk scores, family history, ethnicity or otherwise), it would be useful to compare the cost-effectiveness of using different screening intervals for higher and lower risk men	88.24%	82.35%	0.00%	Consensus reached	Consensus reached	0.93

13. It is useful to compare the cost-effectiveness of using different screening intervals based on PSA level at previous test	76.47%	64.71%	0.00%	Consensus reached	Consensus not reached	0.89

The results show that consensus was considered reached, under both the pre-agreed rules and the more stringent rules, on the usefulness of comparing the following screening strategies in a cost-effectiveness model:

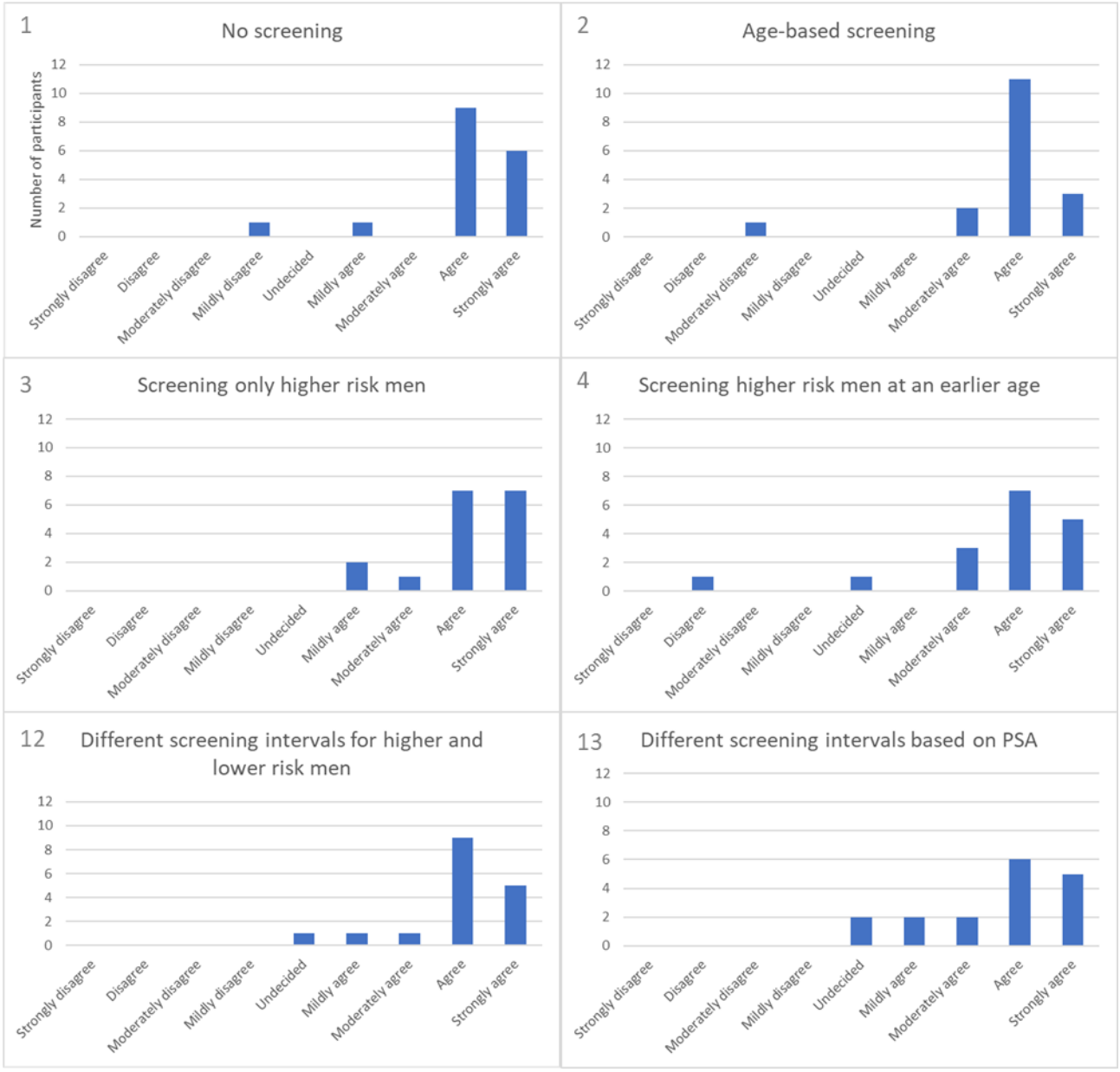
1. No screening
2. Inviting all men within a certain age range to be screened
3. Inviting only higher risk men for screening (if it is possible to identify higher risk men through the use of polygenic risk scores, family history, ethnicity or otherwise)
4. Inviting all men within a certain age bracket for screening but screening higher risk men at an earlier age
5. Using different screening intervals for higher and lower risk men

Consensus was also considered reached under the pre-agreed rules on the usefulness of:

6. Using different screening intervals based on PSA level at previous test

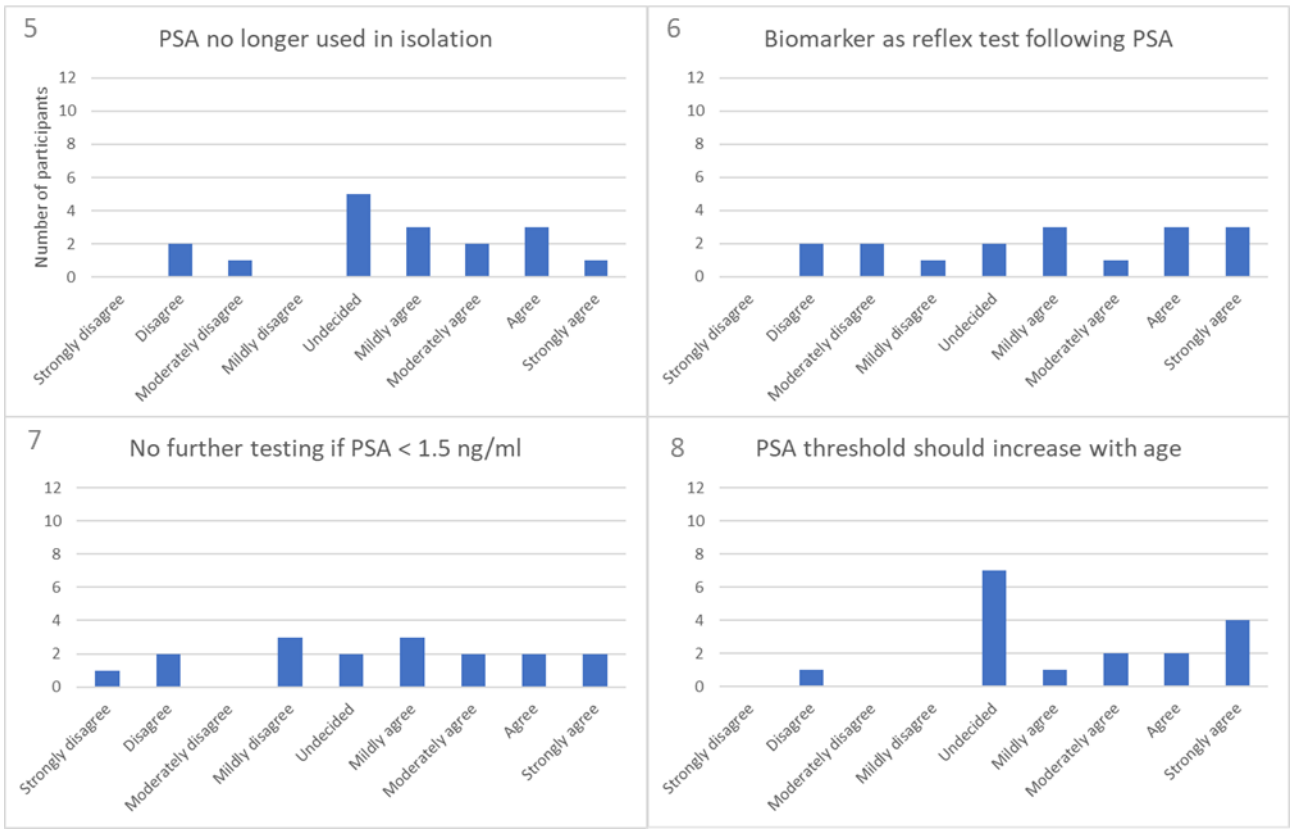
For all of these strategies, interrater agreement was considered high (>0.8).
A detailed breakdown of these responses is shown in Figure 10.

Figure 10. Responses to statements on which consensus was reached (1-4, 12 & 13)



Consensus was not considered reached on aspects relating to PSA and other biomarker testing. The responses suggest that participants did not agree on whether PSA in isolation should be used a reflex test, whether a PSA test should be used alongside a more sophisticated biomarker or risk model, whether a PSA threshold of 1.5 ng/ml has enough negative predictive value to exclude any further testing or whether the threshold for further investigation should increase as men age. In all four of these statements over 35% of respondents indicated agreement but at least 5% indicated disagreement (Figure 11). Comments mainly addressed the need for additional and reliable evidence. There was also uncertainty as to the added benefit of biomarkers over the use of MRI.

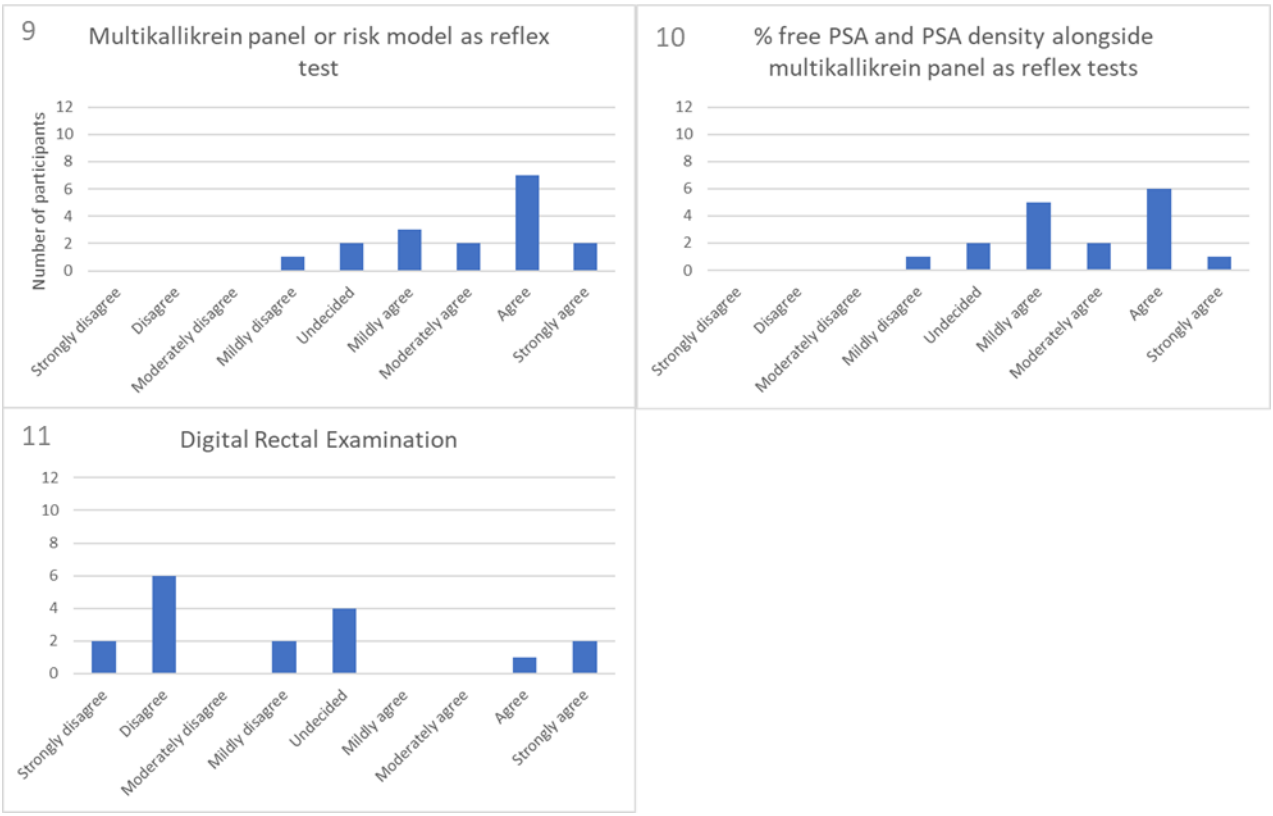
Figure 11. Responses to statements 5-8 showing lack of consensus



Although consensus was not considered reached on whether it would be useful to assess the cost-effectiveness of using reflex tests such as multi-kallikrein panels, risk models, PSA density or % free PSA to triage patients suitable for mpMRI prior to biopsy, there was a tendency towards agreement rather than disagreement with no participants rating either of these statements 1-3. Concerns raised in the comments again centred around a lack of

reliable evidence. With regard to the question of whether all men being screened should be offered a Digital Rectal Examination (DRE), there was a clear tendency towards disagreement with only 18% of participants being in agreement and 47% disagreeing. Concerns raised with offering all men a DRE included no proven utility, increased costs, deterring patients and a high false positive rate (Figure 12).

Figure 12. *Responses to statements 9 – 11 showing lack of consensus*



4.4. Discussion

The aim of this chapter was to illustrate a method, using prostate cancer screening as an exemplar, to identify strategies to be evaluated in cost-effectiveness modelling when there is considerable uncertainty surrounding the relevant comparators. This process has highlighted the uncertainty and diverging views that can exist and a means to focus these views. Views have been elicited from experts working in the field, in accordance with guidelines, and to ensure the findings are applicable to decision makers. Overall agreement was obtained on the patient characteristics and screening technologies to consider in cost-effectiveness modelling. Although the panel did not reach consensus on exact age ranges to screen or a specific screening interval, the cost-effectiveness model developed as part of this dissertation will explore different screening stopping/starting ages and intervals within the limits suggested by the participants.

4.4.1. Comparison with previous work

Husbands et al[251] identified two papers, Sullivan and Payne[262] and Iglesias et al[263], which suggest that the Delphi process could be used to define the boundaries of a model, in model conceptualisation, and to identify face validity. However, this is the first study to both use and illustrate how to use a modified Delphi method to handle decision question uncertainty and identify relevant comparators for a cost-effectiveness analysis in a rapidly evolving decision-making context.

4.4.2. Strengths and limitations

One strength of the study was the panel of experts, who had a wide range of experience. The stability of the panel was good with only 15% attrition. Relying solely on one project team to identify relevant strategies may have resulted in biased views. The modified Delphi method provided a systematic way to gain consensus (according to the predefined criteria) from a wide variety of experts. The web-based format enabled the inclusion of views from respondents from geographically dispersed locations. The anonymity meant that no one voice was given precedence and experts had time to consider their responses. An additional advantage was the relative speed of the process, ensuring relevancy in the context of newer

innovations. A further strength is that the findings from the Delphi can be used to focus future research to provide evidence on aspects of the identified screening strategies that experts feel are important.

A limitation of the web-based format was the inability to have an in-depth discussion with respondents on the meaning and reasoning behind their answers, as would be possible in a face-to-face interview setting, although in most cases the participants provided substantial comments in the free text boxes which helped to explain their decisions. A further limitation is that the snowball approach may have led to researchers only recommending others they agree with, reinforcing any bias in the initial sample. There is also a concern that relevant potentially effective and cost-effective alternatives may have been missed or deliberately excluded by experts if they were not options that they themselves would support. This was mitigated by sending an evidence dossier to participants in advance to ensure they were aware of any relevant evidence and also by selecting participants with varying expertise and from a variety of institutions and countries to encourage diversity of opinions. It is acknowledged, however, that there is still a potential risk of missing relevant aspects when using expert opinion rather than empirical evidence.

The choice to ask participants to indicate their preferred strategy in round 1 did not allow an estimation of numerical uncertainty in their answers. However, participants were encouraged to comment which is where any uncertainty was made clear. In the second round, participants' uncertainty could be more clearly drawn out as they were requested to rank their agreement with the statements on a 9-point Likert scale (strongly disagree to strongly agree).

A further potential limitation is that the study aimed to achieve consensus on relevant screening strategies from participants from different countries with different health care systems. In general, the current standard of care in a particular health care setting may influence the set of comparators chosen which could make incorporating an international perspective difficult. This is less of an issue in prostate cancer screening as current practice in the UK, Europe and the US is a generally consistent policy of shared decision making

around whether or not a man undergoes PSA testing with a recommendation against formal screening. The limitations of an international perspective should be weighed against the benefits in any future applications, however.

4.5. Conclusions

This process has identified the screening strategies that experts considered to be important to compare in a cost-effectiveness model, namely no screening, age-based screening and different risk-stratified approaches incorporating MRI. Although this is not a commonly used approach to identify comparators in the health economic modelling literature, it is one that has proved useful as it has clarified areas of agreement as well as disagreement and established the groundwork for future research. The work has demonstrated that this should be a method that is considered when the decision space is uncertain and rapidly changing, as is the case in prostate cancer screening.

CHAPTER 5. ADAPTATION AND CALIBRATION OF PROSTATA MODEL TO UK SETTING

5.1. Introduction

Following from the modified-Delphi consensus process described in Chapter 4, one of the aims of this dissertation was to assess the impact of the relevant screening strategies identified on the long-term costs, health outcomes and cost-effectiveness of prostate cancer screening in the UK. This was achieved using the Prostata microsimulation natural history model[264] (described in section 5.2), identified in the systematic review described in Chapter 3 as a comprehensive and detailed model, with R code openly available on Github.

This chapter describes the adaptation and calibration of the Prostata model to the UK setting. The model was originally based on a US model[265] and previously used to compare prostate cancer screening strategies in a Swedish setting. The chapter will first describe the natural history model used before detailing how it was: (1) calibrated to the UK context using data from the UK Office of National Statistics (ONS) and 10-year follow-up of the CAP trial (described in section 2.2.8); (2) validated using two large randomised screening trials[21 23]; (3) adapted to reflect screening strategies with stratification by polygenic risk score; and (4) updated to reflect UK parameters. The calibrated model was used to project the lifetime cost-effectiveness of the screening strategies. This is described in Chapter 6.

5.2. Methods

5.2.1. Natural History model

The Prostata model used in this analysis is a continuous time discrete event simulation for the natural history of prostate cancer.[266] It is an open-source model which allows for individual heterogeneity in natural history including disease onset, progression, diagnosis and death. The model is an adaptation of a model developed by the Fred Hutchinson Cancer Research Centre in the US (the PSAPC model) [265 267 268]and allows prostate cancer onset and progression to correlate with change in PSA over time. Modelling change in PSA alongside disease progression is useful as PSA is an observed biomarker that allows inferences to be made about underlying, often unobserved, disease progression. In developing the PSAPC model, Gulati et al first estimated PSA growth and then, conditional on the estimated PSA growth curves, disease progression parameters were estimated.[265] The PSA growth model was informed by data from the control arm of the Prostate Cancer Prevention Trial (PCPT), where 9,000 men were screened for up to 7 years with an exit biopsy regardless of PSA test results.[269 270]

5.2.2. Model structure

The model starts with a cohort of men who are prostate cancer-free at age 35 years but have a specified risk of developing prostate cancer. Prostate cancer onset is modelled via a time-dependent hazard (from age 35) following a Weibull distribution. A man's PSA level is assumed to rise linearly (on the log scale) with age. The slope is higher on average after onset of a low-grade tumour and higher again after onset of a high-grade tumour.

Cancer states and transitions between them

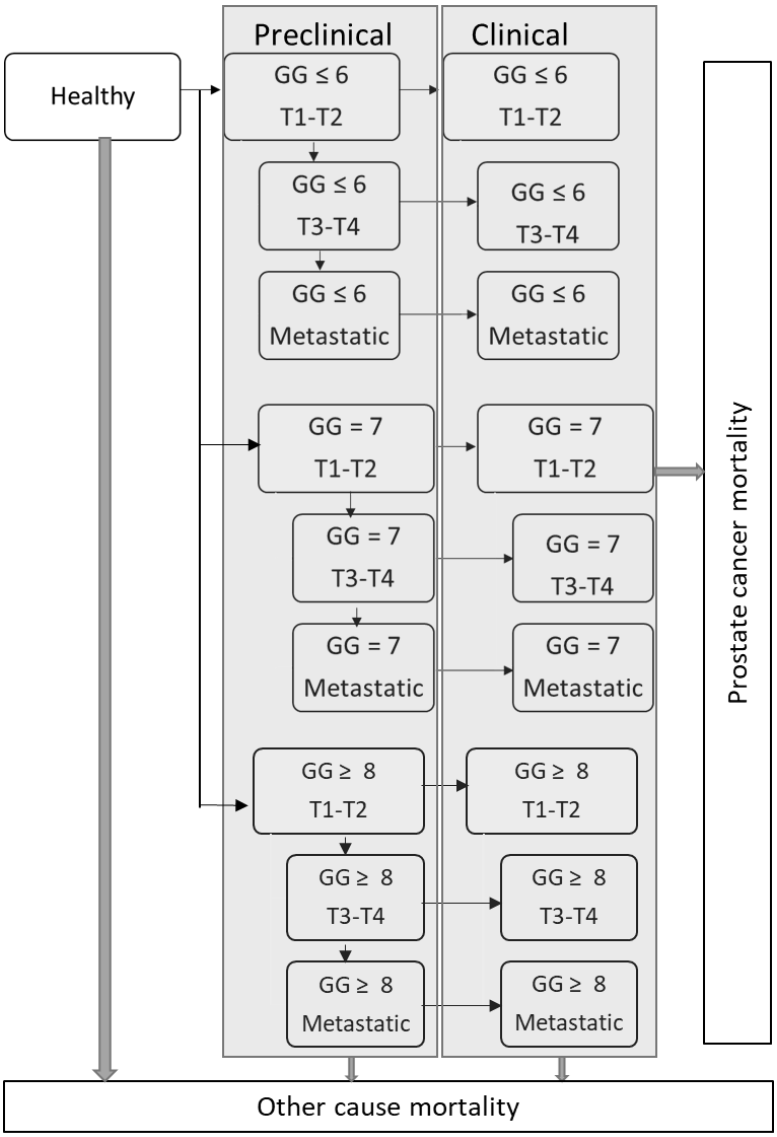
If a man develops prostate cancer, then a Gleason grade (≤6, 7, or ≥8) and cancer stage (T1-2, T3-4, M1) are assigned at cancer onset. A Gleason grade of 6 or less is considered a low-grade cancer, 7 is medium-grade, and 8 or more is high-grade. A lower-grade cancer grows more slowly and is less likely to spread than a high-grade cancer.[43 44] The model assigns Gleason scores dependent on age, with older men more likely to have higher-grade disease.

The model does not include transitions between Gleason scores, with research to suggest that this is a realistic assumption.[271 272]

Progression between cancer stages is modelled and the risk of progression is assumed to increase as a man's PSA increases, as well as his age. This assumption is based on data from the PCPT trial.[269 270] The association between PSA and progression has also been shown in other studies.[273-275] The transition from T1-T2 to T3-T4 is assumed to be the same for all Gleason categories. After onset, the hazard of metastasis follows a Gompertz distribution and is dependent on a man's PSA.

Both preclinical and clinical states are modelled. Preclinical states are defined as men with prostate cancer that is asymptomatic and so has not been detected but could potentially be detected by PSA screening. Once a cancer is detected the man moves to the clinical states. Cancers may be detected either through screening or clinically (any non-screening detection) (Figure 13). A man is more likely to be clinically diagnosed as his PSA increases. The hazard of clinical diagnosis is also assumed to be higher when an individual has metastatic cancer as they are more likely to present with symptoms.

Figure 13. Natural history model diagram



**GG: Gleason Grade*

Survival

The model takes account of both prostate cancer specific and other-cause survival. Baseline prostate cancer survival in the absence of treatment is stratified by cancer stage, Gleason grade, PSA (<10, ≥10) and ten-year age groups. This stratification was previously calibrated to 10- and 15-year survival from the Prostate Cancer data Base Sweden (PCBaSe) database including 93,014 Swedish men diagnosed in the period 1998–2014. Survival for untreated cases is inflated by hazard ratios to obtain survival for treated cases. The treatments modelled are radiation therapy, prostatectomy and conservative management. These are

the primary treatments received by patients in the UK and those recommended in the 2019 NICE prostate cancer guideline.[157 276]

The mechanism of benefit of prostate cancer screening is assumed to be cancer stage-shift, in that the benefit associated with screening is due to a shift to a less advanced cancer stage at diagnosis, resulting in earlier diagnosis and improved survival through potentially curative treatment.

5.2.3. Why this model was chosen

The systematic review described in Chapter 3 identified several prostate cancer natural history models that could have been updated to a UK setting. Only two of these were continuous time discrete-event microsimulation models. As described in the background chapter, a discrete event simulation enables a more accurate reflection of clinical pathways, allowing time-dependent event rates, and avoiding the need for transition probabilities and fixed time cycles, as in a Markov model. This is important in prostate cancer screening as substantial heterogeneity exists in how cancers progress between individuals. The use of a discrete event simulation also facilitates the analysis of risk-stratified and adaptive screening approaches as it allows the simulation of different patient characteristics and diagnostic pathways within the one screening strategy.[110 277]

The other discrete event simulation identified was the MIcrosimulation SCreening Analysis (MISCAN) model used by Heijnsdijk et al.[148] This is a similar model representing the natural history of prostate cancer which has been re-used in different settings and with different decision questions in the past.[13 278-280] Both models consider progression between clinical and preclinical states. In the MISCAN model, cancer progresses through both stages and grades (as opposed to the Prostata model where grade is fixed), but unlike the Prostata model, progression rates are not explicitly correlated with PSA levels. This means that in the Prostata model patients with slow growth in PSA are likely to progress slowly whereas in the MISCAN model patients with slow growth in PSA are as likely to progress quickly or slowly. Previous analyses have shown that the models give slightly different results in terms of prostate cancer incidence and overdiagnosis due to this difference.[279 281 282]

Heijnsdijk et al explain how overdiagnoses may be overestimated in the MISCAN model as slowly progressing cancers that are modelled to have fast PSA growth have the potential to be screen detected and therefore overdiagnosed.[282] In the Prostata model, slowly progressing cancers generally have slow PSA growth and are therefore less likely to be screen detected because they remain below the threshold for further investigation. The model structure, with both stage and grade of cancer incorporated and with progression linked to PSA, was therefore considered to be clinically meaningful. The link between progression and PSA is also important when considering screening strategies where biopsy referral and time to next screen are PSA-dependent.[283]

As mentioned in section 5.1, the Prostata model, rather than the MISCAN model, was also chosen in this case as the code is open-source and easily available on Github (https://github.com/mclements/microsimulation and https://github.com/mclements/prostata).

5.2.4. Summary of changes from earlier models

Several aspects of the earlier Swedish and US models were updated to reflect a UK setting. These included the population structure, PSA re-testing sub-model and distribution of Gleason scores at cancer onset. Other parameters updated included treatment allocation, formal biopsy compliance, background mortality rates and hazard ratios due to prostate cancer treatment. To ensure the model was accurately predicting outcomes in a UK setting, the model predictions were plotted against observed data on prostate cancer incidence and mortality rates. These adaptations will now be described in turn.

5.2.5. Calibration and Validation

Although the prostata model was previously calibrated to a Swedish setting[264], a novel approach was taken to the UK calibration in accordance with the data available. The previous version of the Prostata model used by Karlsson et al[264] and Hao et al[284] calibrated several natural history parameters to Swedish data on the relative distributions of incident cancer stages at diagnoses and prostate cancer survival, and the rate ratio of prostate

cancer incidence from the ERSPC trial. A two-step calibration was used whereby the relative distribution of cancer staging, the mean time from onset to metastatic cancer, and the PSA screening incidence rate ratio were calibrated for first, using a multinomial likelihood and the Nelder-Mead optimisation algorithm. The second step calibrated hazard ratios of survival by age group, cancer stage, Gleason score and PSA values to Swedish survival data.

This section describes the calibration to UK data which took a different approach. The first step involved calibrating the prostate cancer onset parameter and the second the proportion of Gleason grade ≤6, 7, or ≥ 8 cancers by age, study year and study arm at diagnosis. Both steps used a Poisson likelihood and the bound optimization by quadratic approximation (BOBYQA) algorithm. The calibration will be described in the following sections.

Calibration data

The model was first calibrated to a UK setting using national prostate cancer incidence data from the ONS by age[285], and data from the CAP trial on prostate cancer incidence by age and Gleason grade.[23] The ONS publish annual cancer registration statistics for England which detail age-standardised incidence rates for prostate cancer by age. As this dissertation was carried out alongside the CAP trial, access was available to individual level detail from the 10-year follow-up of the trial. Data from the CAP trial is a relevant resource from which to inform the UK-based modelling of prostate cancer natural history, as the trial has followed over 400,000 men since the year 2000 to measure key clinical outcomes. It therefore provides individual level data to inform the rate of cancer progression broken down by Gleason grade, a fundamental component of the natural history model.

Updating model to reflect UK PSA testing

Calibrating the natural history model to data from the CAP trial required accurately modelling organised and background PSA testing in the UK to reflect the levels of such testing in the population enrolled in the CAP trial. The model was therefore adapted to reflect PSA test uptake in the CAP trial by age group, and background re-testing in the UK. Table 9 shows the numbers randomised to the unscreened and screened arms in the CAP

trial by age bracket and the number of those randomized in the screened arm who accepted and received a valid PSA test. The modelled population was first updated to reflect these parameters.

Table 9. *Ages of those randomized to CAP study arms*

Age (yrs)	N randomized to unscreened arm (%)	N randomized to screened arm (%)	N randomized to screened arm who accepted PSA test (%)
50-54	63,423 (28.9)	55,229 (29.2)	17,869 (32.4)
55-59	63,285 (28.8)	55,077 (29.1)	19,448 (35.3)
60-64	51,507 (23.5)	44,057 (23.3)	15,783 (35.8)
65-69	41,224 (18.8)	35,023 (18.5)	11,335 (32.3)
Total	219,439 (100.0)	189,386 (100,0)	64,435 (34.0)

For re-testing, the parameters were estimated using a Weibull cure model (as in the previous version of the Prostata model), stratified by five-year age groups and PSA values at the previous PSA test. These estimates were based on data from Young et al.'s 2010 paper on the number of men retested in the UK within 1 year of their first PSA test (%) (*Table 10*).[286] This study used data on 450,000 men from the Clinical Practice Research Database (CPRD), a large UK primary care database.[287]

Table 10. Data on probability of retesting by age and PSA level

Age group	PSA level	Number retested within 1 year of their first PSA test (%)
50-54	PSA <3	1580/17566 (9%)
	3≤PSA<4	256/948 (27%)
	4≤PSA<6	507/831 (61%)
	6≤PSA<10	336/501 (67%)
	10≤PSA<20	185/294 (63%)
	PSA ≥ 20	82/167 (49%)
55 - 59	PSA <3	2263/20573 (11%)

	3≤PSA<4	336/16800 (2%)
	4≤PSA<6	902/1640 (55%)
	6≤PSA<10	681/1098 (62%)
	10≤PSA<20	315/543 (58%)
	PSA ≥ 20	163/2329 (47%)
60 - 64	PSA <3	2148/16523 (13%)
	3≤PSA<4	308/1711 (18%)
	4≤PSA<6	927/18540 (5%)
	6≤PSA<10	778/1341 (58%)
	10≤PSA<20	408/716 (57%)
	PSA ≥ 20	215/512 (42%)
65 - 69	PSA <3	1992/13280 (15%)
	3≤PSA<4	359/17950 (2%)
	4≤PSA<6	819/20475 (4%)
	6≤PSA<10	934/1639 (57%)
	10≤PSA<20	489/923 (53%)
	PSA ≥ 20	235/618 (38%)

Calibration methods

The rate of prostate cancer onset in the model (γ_o) was first estimated by calibrating the onset parameters to 2017 data on prostate cancer incidence by age provided by the ONS (*Table 11*).

Table 11. ONS data on prostate cancer incidence by age group in the UK in 2017

Age group (yrs)	Rate per 100,000 population
50-54	73.8
55-59	188.7
60-64	339.1
65-69	579.4
70-74	713.9
75-79	810.0
80-84	677.5
85-89	666.4

90 and over	663.2

The calibration used a Poisson likelihood.[288] The formula for the log Poisson likelihood, comparing observed and expected incidence rates by age at diagnosis, was as follows:

$$\iota(\gamma_o) = \sum_i \left\{ O_i \left(1 + log \frac{E_i}{O_i}\right) - E_i \right\},$$

where O_i denotes observed (CAP) number of cases, $E_i = f(\gamma_o)$ denotes expected (Prostata) number of cases as a function of the disease progression parameters, and *i* indexes age groups 50-54, 55-59, 60-64, 65-69.

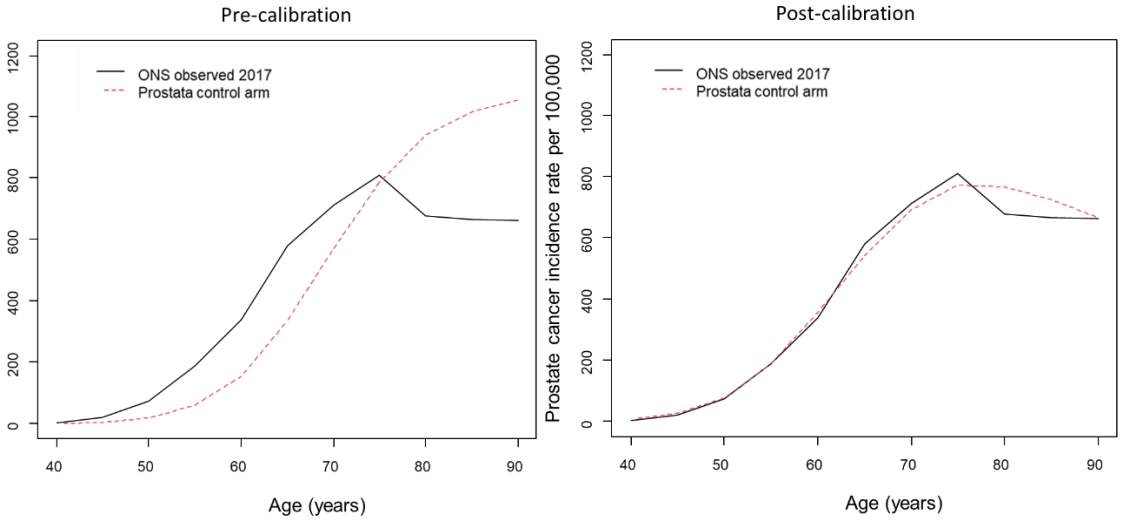
The likelihood compared observed and expected prostate cancer incidence rates by age at diagnosis. Expected prostate cancer incidence rates were based on the sum of the screen and clinical detections produced by the control arm of the model. Observed counts were the corresponding ONS values. As in previous calibrations of the PSAPC model[264 289], and to reduce noise, all other natural history parameters remained fixed.

In the second step, fixing the estimated onset parameters (as well as all other natural history parameters), the proportion of Gleason grade ≤6, 7, or ≥ 8 cancers by age, study year and study arm at diagnosis were calibrated to CAP data after 10 years of follow-up using a separate Poisson likelihood. The likelihood compared observed and expected prostate cancer incidence rates by age, study year, study arm and Gleason grade at diagnosis. Expected prostate cancer incidence rates were based on the sum of the screen and clinical detections produced by model. Observed counts were the corresponding values in the CAP trial. The calibration was programmed in R and used the BOBYQA algorithm.[290] BOBYQA uses finite differences to calculate the derivatives of the optimization target and guide exploration of the sample space.

Calibration results

Figure 14 shows the results of the calibration to ONS data. This highlights the improved fit of the model to the data following calibration, with the incidence rate increased in the younger age groups and decreased in the older age groups to match that seen in the UK data.

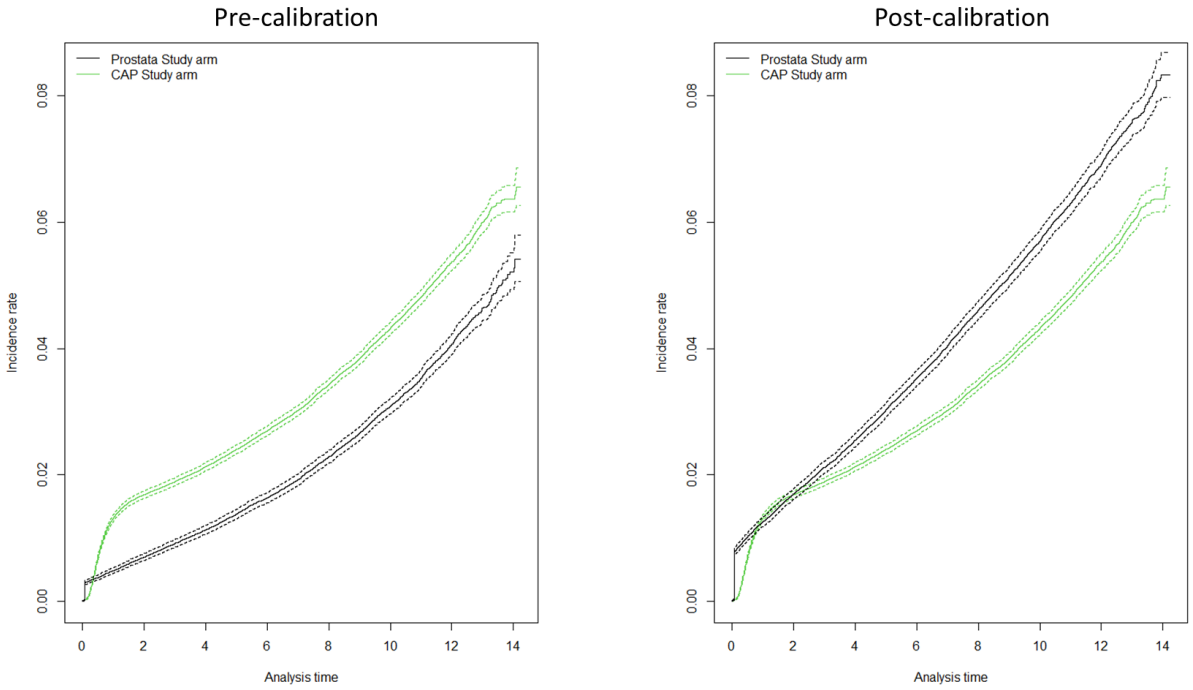
*Figure 14. Prostate cancer incidence by age estimated by the Prostata model compared to observed ONS data – Pre- and post-calibration**



** Estimates of uncertainty not available from ONS data.*

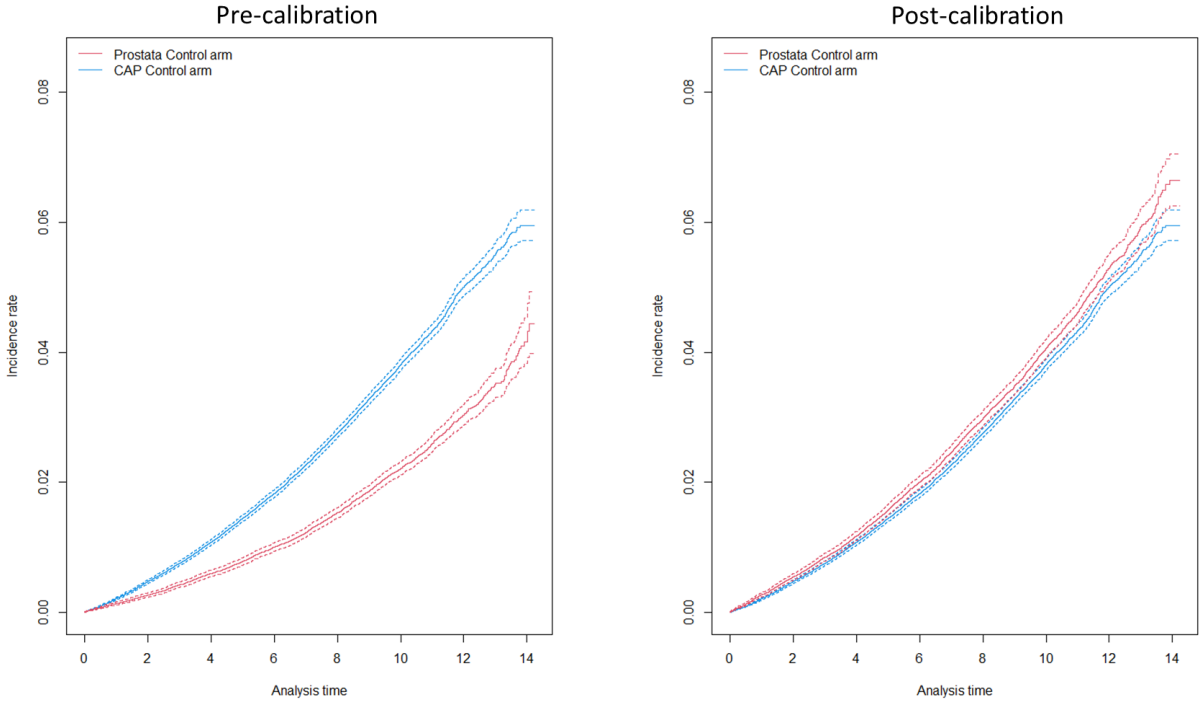
The calibration also improved the model predictions of incidence compared to that seen in the CAP trial (Figure 15 and Figure 16). Prior to the calibration, the model was underpredicting prostate cancer incidence in both the study and control arms at all time points. Following calibration, incidence in the control arm was well predicted. Incidence in the study arm was well predicted at the beginning of the trial but overpredicted following year 3. This is likely related to the large amount of men diagnosed towards the beginning of the CAP trial, with a significant levelling off of such diagnoses in later years. The model is predicting well this initial influx of diagnoses during the PSA screening phase but this adjustment also increases the rate of diagnoses in later years. Methods trialled to improve the calibration are discussed in section 5.2.5.5, however none succeeded in replicating this attenuation of screening effect.

*Figure 15. Cumulative incidence rates by study time in the Study (intervention) arm estimated by the Prostata model compared to observed CAP data**



** Dashed lines show 95% confidence intervals around mean (solid lines).*

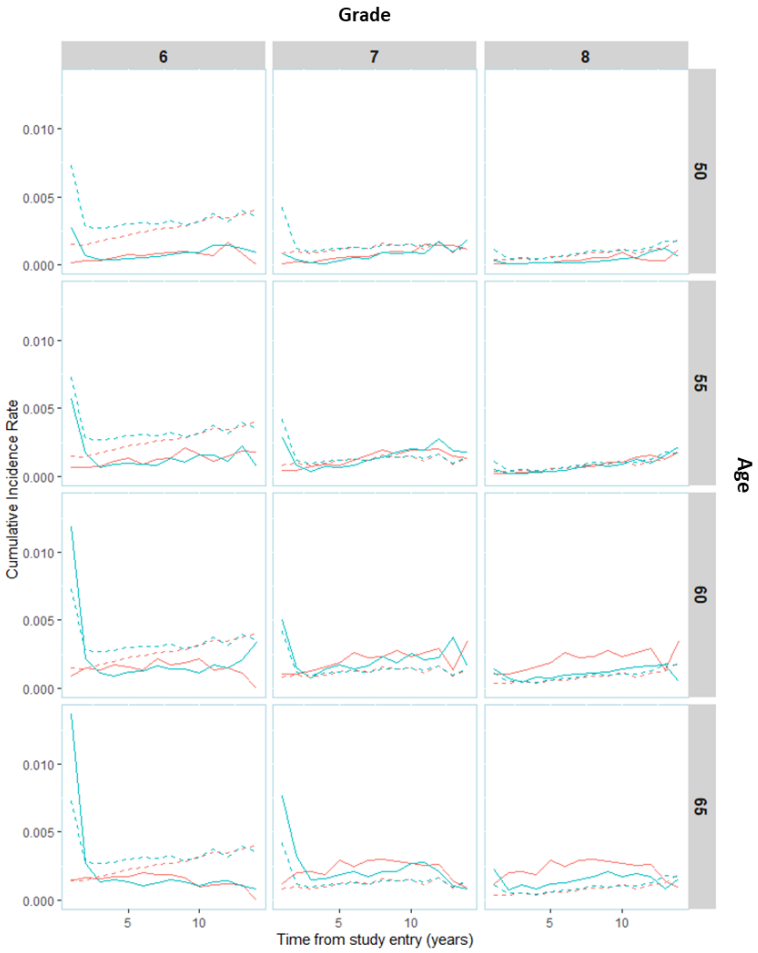
*Figure 16. Cumulative incidence rates by study time in the Control arm estimated by the Prostata model compared to observed CAP data**



** Dashed lines show 95% confidence intervals around mean (solid lines).*

Figure 17 shows the post-calibration plot comparing observed (CAP) and predicted cumulative incidence rates over study time by age at study entry and Gleason grade. This shows that incidence rates were well predicted in most sub-groups. Overprediction was highest in the Gleason grade 6 and below cancers. This large influx of cancers diagnosed towards the beginning of the CAP trial were primarily Gleason grade 6 and below which may be why the model is compensating by consistently overpredicting these cancers. An attempt to target the calibration by down-weighting data on Gleason grade 7 and 8 and above cancers in the likelihood, therefore targeting the calibration at Gleason grade 6 and below, did not improve the fit.

Figure 17. *Post calibration plot comparing observed (full lines) and predicted (dashed lines) cumulative incidence rates by age at study entry and Gleason grade. The red lines relate to the control arm and the blue lines relate to the study arm.*



Other calibration methods explored

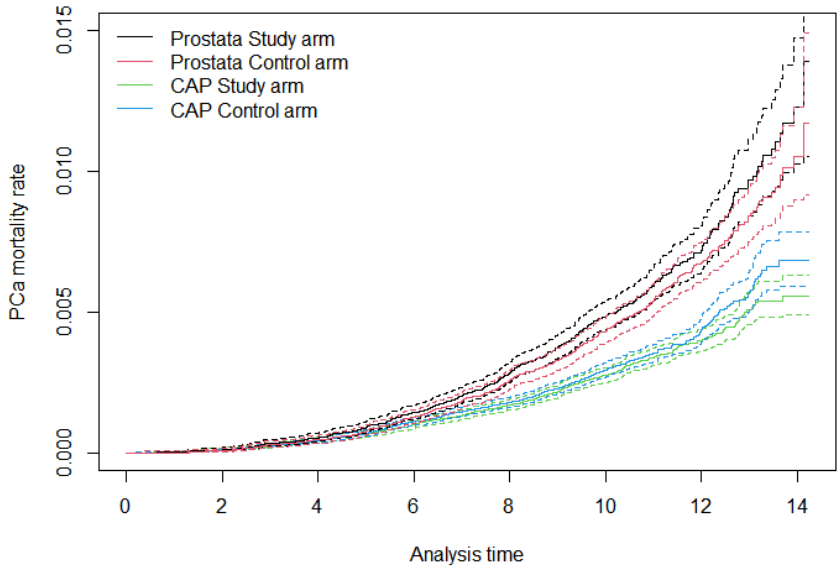
Various alternative methods were explored to improve the fit of the calibration. These included calibrating the rate of clinical diagnosis parameter rather than the rate of onset parameter, calibrating to data on proportions of cases diagnosed within grade groupings and case counts, rather than rates, and calibrating to the control arm of the CAP trial only, rather than to both the study and control arm. Down-weighting data in the likelihood for which the fit was good, therefore targeting the calibration at the less well-fitting areas (e.g. Gleason grade 6 and below cancers), was also trialed. Finally, the impact of using alternative functions in R in place of the BOBYQA algorithm e.g. the optim function with the Nelder-Mead and Brent algorithms was tested[291], as well as using increased simulations. None of these methods succeeded in improving the fit of the calibration, with minimal changes in the predicted incidence rates found.

Validation

To validate the natural history model, prostate cancer mortality predictions were compared to observed ten-year follow-up data from the CAP trial.[23] Prostate cancer mortality rate ratios comparing screening strategies from the ERSPC trial (16 years follow-up with four-yearly screening from either age 55 or age 60 years)[65] and CAP (10 years follow-up after a single screen at age 50 or 60 years) with no screening were also predicted.

For the validation compared with ERSPC, mortality rate ratios over 16 years follow-up of 0.77 and 0.82 for four-yearly screening from ages 55 and 60 years, respectively, compared with no screening were predicted. The observed mortality rate ratio from ERSPC was 0.80 (95% CI: 0.72, 0.89). Compared with CAP, a mortality rate ratio over 10 years follow-up of 0.91 for a single screen at age 50 and age 60 years compared with no screening was predicted. The adherence-adjusted mortality rate ratio from CAP was 0.93 (95% CI: 0.67, 1.29). The predicted prostate cancer mortality was higher in the later years compared with the mortality rates from the CAP trial (Figure 18). Prostate cancer mortality was previously calibrated to a Swedish database[264] which may explain this deviation.

*Figure 18. Validation plot comparing observed (CAP) and predicted (Prostata) cumulative prostate cancer mortality rates by study time**



** Dashed lines show 95% CI around mean (solid lines).*

5.2.6. Modelling polygenic risk

The perceived relevance of risk-stratified screening programmes was a key outcome of the consensus process described in Chapter 4. A further change to the model therefore related to the modelling of polygenic-risk-stratified screening strategies, which were not considered in previous applications of the Prostata model. Polygenic risk of prostate cancer relates to how different sources of genetic variation influence disease risk.[77] Men with a high polygenic risk score are known to be more susceptible to prostate cancer than men with a low risk. If the polygenic risk is known and varies across a population, then this offers the potential of providing risk-stratified screening programmes. Several recent cost-effectiveness analyses have considered polygenic-risk-stratified screening in prostate cancer.[28 292 293] Two of these have been UK based but have used a life-table to model long-term consequences of prostate cancer screening, considering only time to onset of prostate cancer and mortality, rather than a detailed natural history model accounting for cancer progression.[28 293] The same basic assumptions with regard to risk-stratified screening as made in the recent UK papers were applied in updating this model.

Pharoah et al show that the distribution of polygenic risk on a relative risk scale in the population at birth is log-normal.[294] A normal distribution is defined by its mean value (μ) and its standard deviation (σ). For a log-normal distribution, the mean is set so that the arithmetical average risk is equal to 1. This is achieved by setting $\mu = -\sigma^2/2$. The variation in risks in the population, and therefore the ability to identify individuals as high or low risk, is defined by the standard deviation. Pharoah et al also show that the risk distribution in cases is log-normal, as in the general population, but shifted (on a log-scale) by σ^2.
It was assumed that known prostate cancer susceptibility genetic variability follow a log-normal distribution with mean -0.68/2 and variance 0.68 on the natural logarithm scale, such that the frailty has mean 1.[28 50 53] This is based on the 175 susceptibility loci for prostate cancer that have been identified in genome-wide association studies.[52] It was additionally assumed that there was unmeasured genetic variability on a log-normal distribution with a mean of -1.14/2 and a variance of 1.14, where again the frailty has a mean of 1. This was based on evidence from Kiciński et al who carried out a meta-analysis of 33 studies reporting on the impact of a family history of prostate cancer on disease incidence.[49]

Estimates of 10-year absolute risk of a prostate cancer diagnosis for yearly age groups from 50 to 69 were based on Callender et al[293] who calculated these using the mean of the incidence of prostate cancer, mortality from prostate cancer and mortality from other causes between 2013 and 2016 as recorded by the Office for National Statistics (*Table 12*).[293] From these, an age-specific rate ratio for the 10-year absolute prostate cancer risk of 7.5% compared with the population risk was derived using the formula:

$$rr = log(1-0.075) / log(1- 10\text{-year risk})$$

Polygenic-risk-stratified screening was implemented in the model such that men with a genetic variability above the rate ratio were eligible for screening. This probability was calculated using the formula:

$$1-\Phi (log(rr), \mu, \sigma 2)$$

where Φ is the cumulative distribution function of the standard normal distribution.

Table 12 shows the resulting estimated probabilities of being above the 7.5% risk threshold at each age and the proportion of cases above the threshold.

Table 12. Population 10-year risk of developing prostate cancer, probability above 7.5% risk threshold and proportion of cases above threshold by age

Age (yrs)	10-year risk	Probability above 7.5% risk threshold	Proportion of cases above threshold
50	0.013	0.005	0.040
51	0.015	0.008	0.057
52	0.018	0.015	0.088
53	0.02	0.020	0.110
54	0.023	0.030	0.146
55	0.026	0.042	0.183
56	0.03	0.060	0.233
57	0.033	0.076	0.271
58	0.037	0.098	0.320
59	0.041	0.122	0.366
60	0.045	0.147	0.410
61	0.049	0.172	0.452
62	0.052	0.192	0.481
63	0.056	0.218	0.518
64	0.059	0.238	0.544
65	0.062	0.257	0.569
66	0.065	0.277	0.592
67	0.067	0.290	0.607
68	0.069	0.302	0.621
69	0.071	0.315	0.634

5.2.7. Updating model parameters

A further step taken to update the model to a UK setting involved identifying data sources to inform other model parameters. These were primarily sourced from the systematic review of previous economic models described in Chapter 3.Firstly, the accuracy of biopsy and pre-biopsy mpMRI was updated based on Hao et al[295] who carried out a meta-analyses using data from 16 cross-sectional studies comparing diagnostic tests identified in the agreement analysis of a 2019 Cochrane review.[296] The review aimed to determine the diagnostic accuracy of MRI, MRI-targeted biopsy and systematic biopsy compared to template-guided biopsy in men with suspected prostate cancer. The data were used to estimate the specificity and sensitivity (Gleason ≤6, or Gleason ≥7) for: an mpMRI result; a systematic biopsy; and mpMRI targeted biopsies given a positive pre-biopsy MRI. The estimates are shown in Table 13.

No alternative estimates of diagnostic accuracy to fit the parameters in the model could be sourced from the literature. Of the studies included in the systematic review described in Chapter 3, the 2021 prostate cancer model developed for the NICE guideline[157] estimated the sensitivity of mpMRI targeted biopsies by MRI Likert score and whether the cancer was clinically significant. These were based on data from the PROMIS trial[25] and the clinical review carried out as part of the guideline which mainly included data from the PRECISION trial[78]. Neither of these trials were included in the agreement analysis of the Cochrane review: PROMIS as it compared MRI to template-guided biopsy rather than systematic biopsy and PRECISION as it did not perform the index tests and/or reference standard in the same men. The NICE report did not provide a breakdown of diagnostic accuracy by Gleason grade.

Faria et al[152] provided estimates of the diagnostic performance of systematic biopsy based on individual patient data from the PROMIS trial. Similarly, the breakdown was provided by low (men with Gleason score < 7 and PSA<10), intermediate (men with Gleason score=7 or 10≤PSA<20) and high-risk cancer (men with Gleason score ≥8), and so did not fit the parameters of the Prostata model. Callender et al[28] also used the Cochrane review to obtain estimates of the accuracy of MRI.

Table 13. *Accuracy parameters sourced from Hao et al*[297]

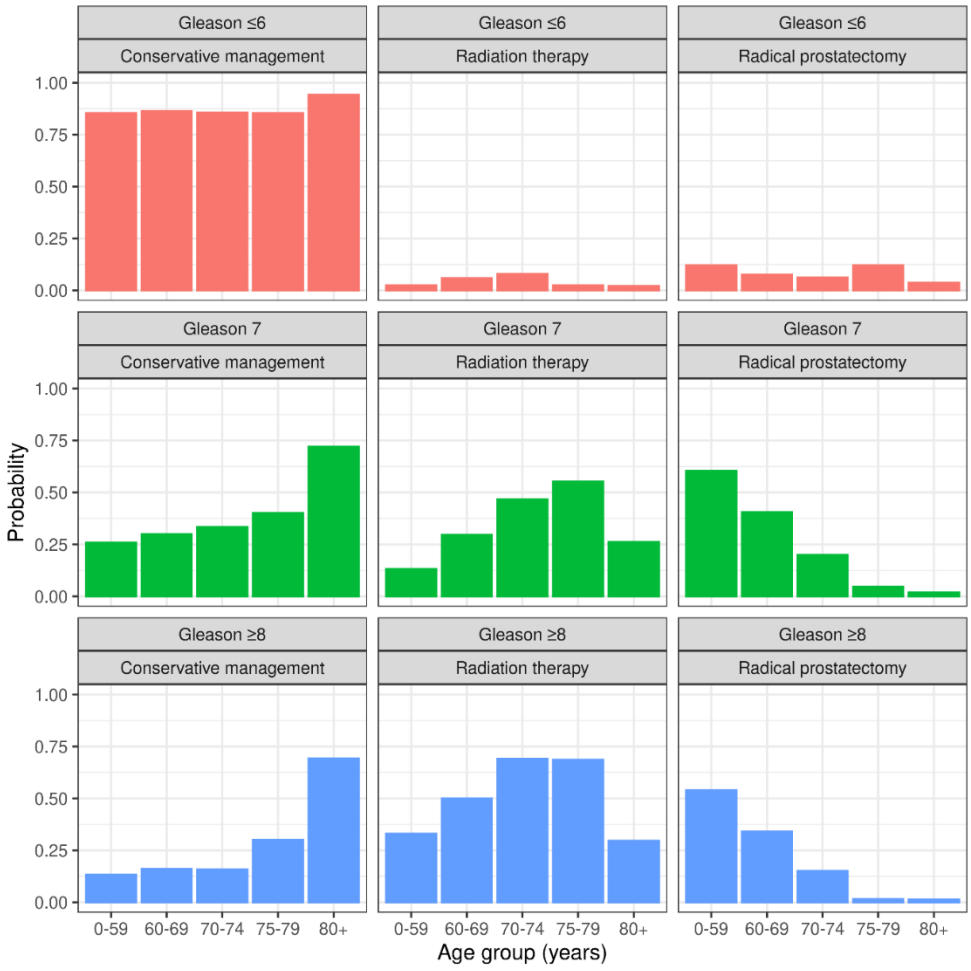
Parameter	Estimate	95% CI
Specificity for mpMRI = Pr(mpMRI-|Healthy)	0.548	(0.435, 0.657)
Probability of positive mpMRI results (Healthy)	0.452	(0.343, 0.565)
Sensitivity for mpMRI (GG <= 6)	0.715	(0.614, 0.798)
Sensitivity for mpMRI (GG >= 7)	0.931	(0.893, 0.956)
Sensitivity for standard biopsy (GG <=6)	0.860	(0.824, 0.889)
Sensitivity for standard biopsy (GG >=7)	0.897	(0.809, 0.947)
False negative rate of standard biopsy (GG <=6)	0.140	(0.111, 0.176)
False negative rate of standard biopsy (GG >=7)	0.103	(0.053, 0.191)
Sensitivity for mpMRI-targeted biopsy (GG <=6)	0.753	(0.568, 0.875)
Sensitivity for mpMRI-targeted biopsy (GG >=7)	0.934	(0.889, 0.962)
False negative rate of mpMRI-targeted biopsy (GG <=6)	0.247	(0.125, 0.432)
False negative rate of mpMRI-targeted biopsy (GG >=7)	0.066	(0.038, 0.111)

Treatment allocation

Management pathways in the model included treatment by conservative management, radical prostatectomy or radical radiotherapy, post-treatment follow up, palliative therapy and terminal care. Probabilities for treatment assignment to either conservative management, radical prostatectomy or radical radiotherapy were updated based on the most recently available estimates from the National Cancer Registration and Analysis Service (NCRAS, 2016), which provided data on all people living in England who were diagnosed with cancer.[298] Values were stratified by five-year age groups and Gleason grade.

Figure 19, created using NCRAS data, shows that people with lower grade cancer (Gleason grade 6 or less) were generally treated with conservative management (which is a combination of active surveillance and watchful waiting), with few receiving either radiotherapy or radical prostatectomy. In Gleason grade 7 and Gleason grade > 7 cancers the trend was for prostatectomy in the younger age groups, radiotherapy in the 60–79-year-olds and conservative management in those aged 80 years and over.

Figure 19. National Cancer Registration and Analysis Service primary treatment allocation estimates 2016



The studies identified in the systematic review did not report treatment allocation by age group and Gleason grade. A similar trend was identified however with estimates of between 10% and 50% (average 35%), and 10% and 30% (average 17%) assumed for assignment to radical prostatectomy and radiation therapy in low grade cancers, respectively. In high grade cancers, estimates of between 40% and 100% (average 72%), and 25% and 40% (average 29%) were assumed for assignment to radical prostatectomy and radiation therapy, respectively. [145 147 150 151 156 299]

Other parameters

Background mortality was based on UK life tables.[300] The mortality hazard ratio assumed for prostate cancer specific death in radical treatment over conservative management was 0.63

(95% CI, 0.21 to 1.93) based on the ProtecT trial.[301] This was updated from the estimate of 0.56 (95% CI, 0.41 to 0.77) used in the Prostata model based on the mortality hazard ratio for surgery over watchful waiting from the earlier Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4) trial.[213] The SPCG-4 trial assigned 695 men with prostate cancer at centres in Sweden, Finland, and Iceland to either watchful waiting or radical prostatectomy between 1989 and 1999 and followed them to 2012. The estimate from the ProtecT trial is based on 1098 UK men assigned to either active surveillance or radical prostatectomy between 1999 and 2009 and followed for a median of 10 years, making it more recent and representable of the population included in the model.

Biopsy compliance in men with raised PSA was assumed to be 85.3% based on data from CAP.[69] This was very similar to biopsy compliance assumed in the previous version of the model which was based on the ERSPC trial (85.6%).[21]

Costs

Costs are from the perspective of the UK National Health Service (NHS) and correspond to the 2020 price year. The costs of obtaining polygenic risk scores are uncertain. A recent systematic review of cost-effectiveness analyses of screening interventions informed by polygenic risk scores[302] found that the cost assumed for collecting this information in the context of prostate cancer varied from £25[28 293] to €255 (£217) (the latter including the cost of a GP visit).[143] In the absence of a better source, the estimate used in this model (£25) was based on that of the recent UK papers by Callender and colleagues[28 293], who gathered this information "from personal discussion of costs charged to NHS hospitals for prostate cancer genome wide associations studies".

The costs of assessing suspected prostate cancer, prostate biopsy, radical prostatectomy, radical radiotherapy and active surveillance were also taken from Callender et al 2021.[28] These estimates were chosen as they were recent, UK-based and fitted well with the parameters of the model. Other potential sources of information for these parameters were the health economic report from the NICE prostate cancer guideline[157] and the study by

Faria et al[152], however both provided a breakdown of costs by risk group (low, intermediate, high), rather than by treatment.

The cost of a PSA test and mpMRI were based on estimates from the NICE health economics report from the 2019 guideline on prostate cancer diagnosis and management[157], updated to the 2020 financial year. This was done using the CCEMG-EPPI web-based tool.[229] Terminal care costs were based on model-based estimates of the direct health care cost associated with men with prostate cancer at the end of life.[303] The study by Round et al estimated resources used by an individual with prostate cancer in the last period of life from UK Hospital Episode Statistics[304] and combined these with an estimate of the unit cost for each of those resources from NHS reference costs. Cost estimates along with a breakdown of components included in the costs are shown in *Table 14*.

Table 14. Cost parameters

Parameter	Estimate	95% CI	Source
PSA test	£21	(17-25)	NICE guideline, 2021.[157] Based on cost of a PSA test kit and nurse consultation.
Polygenic risk stratification	£25	(20-30)	Callender et al, 2021.[28] Estimated from costs charged to NHS hospitals for prostate cancer genome-wide association studies.
Biopsy (Systematic/MRI-targeted)	£581	(465-697)	Callender et al, 2021.[28] Weighted average of cost of transrectal ultrasound guided and perineal biopsy. Includes relevant histopathology, potential admission for sepsis and cost of a urological appointment.
Multiparametric MRI	£339	(271-407)	NICE guideline, 2021.[157] Includes time of two radiographers, an appointment with a consultant, and equipment, administration and consumable costs.
Assessing suspected prostate cancer	£545	(436-654)	Callender et al, 2021.[28] Includes an isotope bone scan, assessment by a urological multi-

			disciplinary team and a further outpatient urological appointment
Prostatectomy	£9808	(7846-11770)	Callender et al, 2021.[28] Includes an appointment with a urologist and a weighted average of the cost of major open, robotic and laparoscopic radical prostatectomies from NHS reference costs.
Radiation therapy	£6462	(5170-7754)	Callender et al, 2021.[28] Includes an appointment with a clinical oncologist, preparation for Intensity Modulated Radiation Therapy (IMRT), and outpatient delivery of treatment on a megavoltage machine
Active surveillance (yearly)	£577	(462-692)	Callender et al, 2021.[28] Includes cost of 3 PSA tests and 2 urological appointments. Assumes a third of men will need an annual mpMRI and biopsy.
Palliative care/Terminal illness	£7383	(5906-8860)	Round et al, 2015.[303] Model assumed terminal care for the six months prior to a death due to prostate cancer and palliative care for the 6-30 months prior.

Utilities

Health state utility values were sourced primarily from Hao et al and based on estimates obtained using the EQ-5D instrument.[297] As mentioned in section 2.3.2, the EQ-5D measurement method is the preferred method to measure health-related quality of life in the UK based on NICE guidance.[305] The health state values used for a biopsy, a cancer diagnosis and terminal illness were taken from Heijnsdijk et al[190], the source most commonly used in the studies identified in the systematic review in Chapter 3. As mentioned in Chapter 3, the estimate used for a biopsy in Heijnsdijk et al[190] was taken from an earlier study[191] that focused on breast cancer biopsy. The lack of a more appropriate estimate was confirmed by a recent systematic review of health state utility values by Li et

al (2019)[306] who found no disutilities associated with the prostate cancer screening procedure.

The estimate used for a cancer diagnosis was based on a study by Korfage et al who obtained EQ-5D valuations from 52 men participating in the ERSPC trial who completed EQ-5D questionnaires before and after a cancer diagnosis.[192] The utility associated with active surveillance was based on Loeb et al who measured utilities among 37 US men on active surveillance for prostate cancer participating in focus groups between 2015–2016, using the EQ-5D questionnaire.[307] The utility associated with the first two months following radical prostatectomy and radiation therapy was taken from a UK study by Hall et al who collected EQ-5D data from 147 men with prostate cancer being treated at two UK hospitals.[308] The utility associated with the following 10 months following prostatectomy and radiation therapy was based on a meta-analysis of three studies,[308-310] including Korfage et al and Hall et al, carried out by Hao et al.[295] The third study collected EQ-5D data from 411 men who had prostate surgery in 34 UK centres.[310]

The utility associated with metastatic cancer was also taken from a systematic review and meta-analysis of six studies carried out by Hao et al. All studies reporting utility values associated with metastatic cancer in prostate cancer patients up to 2019 were included. The utility associated with the post-recovery period was based on a meta-analysis of two studies; Torvinen et al collected EQ-5D data from 309 prostate cancer patients in Helsinki more than 1.5 years after a diagnosis[187] and Watson et al collected data from 316 men diagnosed with prostate cancer 9–24 months previously at two UK cancer centres.[311] Finally, the utility associated with palliative therapy was based on a meta-analysis of three studies, carried out by Magnus et al[312]. One of these was Torvinen et al who also collected data in 17 patients receiving palliative care. Another was Färkkilä et al who collected EQ-5D data from 30 prostate cancer patients receiving palliative care. This study was also based in Helsinki. The third study, Wu et al, obtained EQ-5D data from 270 prostate cancer patients on palliative care participating in a multinational observational study.[313]

QALY norms for the UK were used to reflect background age-specific quality of life, which decreases naturally with age.[314] All state QALYs were multiplied by their age-specific norms to reflect this natural reduction in quality of life. The values used are shown in *Table 15* along with their method of estimation.

Table 15. Utility parameters and methods of estimation

Parameter	EQ-5D Estimate	95% CI	Source	Method	Participants
Biopsy	0.90	(0.87 - 0.94)	Heijnsdijk 2012[190] sourced from de Haes 1991[191]	Symptoms and functional levels associated with diagnostic phase in breast cancer screening summarized following literature review. Visual analogue scale (VAS) given underneath state description. Respondents asked to mark evaluation of state, with anchors worst (score = 0) and best imaginable (score = 100) quality of life.	15 employees of the Department of Public Health and Social Medicine and 13 experts in breast cancer treatment and epidemiology in the Netherlands
Cancer diagnosis	0.80	(0.75 - 0.85)	Heijnsdijk 2012[190] sourced from Korfage 2006[192]	VAS anchored with worst (score = 0) and best imaginable (score = 100) health state. Participants asked to indicate how good or bad their current health perceived to be before diagnosis, before initiation of treatment, and 6 months afterwards.	52 screen detected prostate cancer patients participating in ERSPC trial
Prostatectomy part 1 (first 2 months)	0.83	(0.73 – 0.91)	Hall 2015[308]	EQ-5D-3L for men receiving surgery collected within 6 months of diagnosis. Preferences for	130 men with prostate cancer being treated at two UK hospitals

				health states elicited from UK general public (Dolan 1995[315])	
Prostatectomy part 2 (next 10 months)	0.89	(0.88 - 0.91)	Meta-anaysis of Glazener 2011[310], Hall 2015[308] and Korfage 2005[316]	Glazener 2011: EQ-5D-3L collected at baseline and 6- and 12-months post-prostatectomy. Hall 2015: EQ-5D-3L for men receiving prostatectomy collected within 9 months and 15 months post-diagnosis. Korfage 2005: EQ-5D collected at 6- and 12-months post-prostatectomy. Preferences for health states elicited from UK general public (Dolan 1997[317])	Glazener 2011: 184 UK men with prostate cancer approached at time of admission for prostate surgery or at pre-operative assessment clinics Hall 2015: 130 men with prostate cancer being treated at two UK hospitals Korfage 2005: 123 men with prostate cancer recruited from 4 hospitals in the Netherlands prior to prostatectomy
Radiation therapy part 1 (first 2 months)	0.82	(0.75 - 0.88)	Hall 2015[308]	EQ-5D-3L for men receiving radiotherapy collected within 6 months of diagnosis. Preferences for health states elicited from UK general public (Dolan 1995[315])	130 men with prostate cancer being treated at two UK hospitals
Radiation therapy part 2 (next 10 months)	0.83	(0.88 - 0.91)	Meta-anaysis of Hall 2015[308] and Korfage 2005[316]	Hall 2015: EQ-5D-3L for men receiving radiotherapy collected within 9 months and 15 months post-diagnosis. Korfage 2005: EQ-5D collected at 6- and 12-months post radiation therapy.	Hall: Described above. Korfage 2005: 187 men with prostate cancer recruited from 4 hospitals in the Netherlands prior to radiation therapy

Active surveillance	0.90	(0.85 - 0.95)	Loeb 2018[307]	EQ-5D-3L completed prior to focus group discussion	37 men with prostate cancer on active surveillance from 2 US hospitals
Post recovery period	0.86	(0.84 - 0.88)	Meta-analysis of Torvinen 2013[187] and Watson 2016[311]	Torvinen 2013: EQ-5D-3L with most commonly used UK time-trade-off tariff (no reference given) Watson 2016: EQ-5D-5L with crosswalk algorithm used to convert 5L to 3L.[318]	Torvinen 2013: 317 men from the Helsinki and Uusimaa Hospital District with local disease, more than 1.5 years after diagnosis Watson 2016: 316 men diagnosed with prostate cancer 9–24 months previously in two UK cancer centres
Palliative therapy	0.62	(0.58 – 0.66)	Magnus 2019[312] (Meta-analysis of Farkkila 2014[319], Torvinen 2013[187], Wu 2007[313])	Farkkila 2014: EQ-5D-3L using UK time-trade off tariff[320] Torvinen 2013: EQ-5D-3L with most commonly used UK time-trade-off tariff (no reference given) Wu 2007: EQ-5D collected at enrollment and 3, 6, and 9 months after enrolment	Farkkila 2014: 28 prostate cancer patients in Helsinki with metastatic disease and receiving palliative treatments only or who died due to cancer within 6 months of responding to the questionnaire Torvinen 2013: 19 men with prostate cancer from the Helsinki and Uusimaa Hospital District receiving palliative care Wu 2007: 280 Metastatic Hormone-Refractory Prostate Cancer Patients from North America, Europe, and Australia.

Terminal illness	0.40	Held constant	Heijnsdijk 2012[190] sourced from Konski 2007[198] (sourced from Bennett 1996[195]), Penson 2005[321] (sourced from Bayoumi 2000[322], sourced from Bennett 1996[195]) and Ramsey 2005[201] (sourced from Bayoumi 2000[322], sourced from Bennett 1996[195])	Bennett 1996: Focus groups followed by time-trade off exercise	Bennett 1996: 23 US urologists and 18 oncologists who treated large numbers of prostate cancer patients.
EQ-5D index population norms 18-24	0.934	Held constant	Janssen and Szende[314] based on Kind 1998[323]	EQ-5D-3L administered via interview	A random sample of 3395 individuals selected from the general UK population
25-34	0.922				
35-44	0.905				
45-54	0.849				
55-64	0.804				
65-74	0.785				
75+	0.734				

5.3. Discussion

In summary, this chapter has described the adaptation of the Prostata model to allow the comparison of screening strategies in a UK setting. The calibrated natural history model has shown good prediction of prostate cancer incidence in the UK when compared to data from the ONS and the control arm of the CAP trial. Incidence in the study arm of the CAP trial is well predicted at the beginning of the trial but overpredicted in later years. The validation exercise gave prostate cancer mortality rate ratios that were broadly consistent with the CAP and ERSPC trials, although the prostate cancer mortality rate was overpredicted in the

later years of the CAP trial. Overall, it could be said that the natural history model is valid and can give reasonable predictions of the impact of introducing screening strategies in the UK. The Delphi consensus process in Chapter 4 highlighted risk-stratified screening as a strategy of relevance. The model has therefore been updated to enable an assessment of the cost-effectiveness of such strategies. The use of UK-specific data on other parameters such as costs, treatment allocation, and mortality has additionally prepared the model for use in a UK setting.

5.3.1. Comparison with previous studies

Previous prostate cancer models using calibration to determine parameters

The calibration and reuse of natural history models is common in the prostate cancer literature. The CISNET prostate working group, including investigators from the Fred Hutchinson Cancer Research Center in the US and the Erasmus University Medical Center in the Netherlands, have used the same core natural history models (PSAPC and MISCAN-PRO) to assess the cost-effectiveness of various prostate cancer screening strategies, in multiple settings.[28 293] The Prostata model used in this analysis, which is based on the PSAPC model, has also been re-used several times for different decision questions including whether the use of MRI and/or the Stockholm3 test is cost-effective in prostate cancer screening.[143 284 295] The calibration process, whereby each model identifies natural history parameters that are most consistent with observed data in the relevant setting, is key.

Several of the other cost-effectiveness models identified in the systematic review in Chapter 3 used calibration to determine model parameters. Faria et al[152] calibrated all transition probabilities in their model (progression free to metastases, progression free to death and metastases to death) to US and UK data on life expectancy by cancer risk group and treatment, the proportion of patients metastasised and the probability of dying after metastasis.[324 325] The calibration model randomly drew numbers from the data available, with several conditions of plausibility, until 1,000 plausible sets of transition probabilities for

each subgroup were found. The cost-effectiveness model for the NICE prostate cancer guideline calibrated progression parameters (low to intermediate risk, intermediate to high-risk and high risk to metastases) to Scandinavian and UK data on incidence of metastases and prostate cancer death.[213 216 326] Numerical optimisation was used to estimate the optimal value of the parameters by minimising the error in the total number of people developing metastases or dying from prostate cancer. This was done using the generalised reduced gradient nonlinear algorithm[327] used by the Solver add-in in the software Excel.

Previous cost-effectiveness models considering polygenic risk-stratified screening

The analysis reported in this chapter has taken the same approach as that of Callender et al[293] to the modelling of polygenic risk, assuming that men begin screening at the age that a certain risk threshold is met. A recent systematic review identified only one other cost-effectiveness analysis considering polygenic-risk-stratified screening in prostate cancer.[302] This study[292], based in the US and using the PSAPC model, compared age-based screening at different age ranges and screening intervals to genetic risk-stratified screening strategies in which men at average risk receive the standard age-based screening policy and low- and high-risk men received lower and higher intensity screening strategies.

5.3.2. Strengths and limitations

This analysis has calibrated a detailed natural history model using individual patient data from a rich UK data source (the CAP trial). This dataset was a valuable resource from which to inform the updated UK-based modelling of the natural history of prostate cancer due to its long-term follow-up and, being set in the UK where prostate cancer mortality is amongst the highest in Europe, its direct relevance to UK policy. Despite not accurately predicting the attenuation of screening effect in later years of the CAP trial, the natural history model provided a good fit on visual inspection to data on prostate cancer incidence from the ONS and control arm of the CAP trial and predicted mortality rate ratios that were close to the point estimates from ERSPC and CAP, and well within their 95% confidence intervals. It is the first UK study to model the impact of polygenic risk scores on prostate cancer onset, considering how those cancers then progress over time. It has also updated other UK-

specific parameters to provide a model well-suited to modelling the impact of contemporary screening strategies in the UK, where uncertainty as to the value of screening exists.

A limitation is the failure of the calibration to accurately predict the attenuation of screening effect in the later years of the CAP trial, as well as incidence in Gleason grade 6 and below cancers. The model is therefore predicting a higher ongoing incidence of lower grade prostate cancer diagnosed as a result of screening than that which was observed in the CAP trial. The model is also overestimating prostate cancer specific mortality in the later years of the CAP trial. Several methods were attempted to improve the calibration although none succeeded in achieving a closer match of model predictions with data. The impact of these deviations on cost-effectiveness analysis may be that the predicted costs associated with screening are higher than in reality, due to the treatment and monitoring of lower grade cancers, and costs associated with prostate cancer deaths. Overdiagnosis rates may also be overestimated. It is difficult to judge the impact this overprediction may have on the relative cost-effectiveness of screening interventions.

In addition, an assumption is made that data from the CAP trial is representative of a UK setting. The trial was initiated in the year 2000 so may not be representative of current clinical practice. Diagnosis and management have now changed with the introduction of MRI and more of a focus on active surveillance of low grade cancers.[157] There is also the possibility that cancers diagnosed in the CAP trial would be staged differently if diagnosed today.[20]

Where possible the model was updated to use UK data. However, several aspects of the previous version of the model were left unchanged such as the longitudinal PSA sub model i.e. how PSA growth is linked with prostate cancer progression. The PSA sub model was informed by data from the control arm of a randomized trial of finasteride for the prevention of prostate cancer, the Prostate Cancer Prevention Trial (PCPT).[269 270] This trial performed an end-of-study biopsy in all men who had not been diagnosed with prostate

cancer on symptomatic presentation. These data enabled the accurate estimation of pre- and post-onset PSA slopes, but no comparable data was identified in the UK.

A further limitation relates to the assumption that conservative management and active surveillance are interchangeable when informing the treatment allocation parameters. Treatment allocation data from NCRAS[298] provided estimates for conservative management, which is a combination of active monitoring and watchful waiting. Active surveillance and watchful waiting can have different effects, however, as the former involves regular testing/biopsies while the latter does not.[328] Other parameters in the model relate to active surveillance specifically including the mortality hazard ratio for radical treatment, which is based on a comparison with active surveillance from the ProtecT trial[301], and the cost and utility parameters assigned to this health state.[28 307]

An additional limitation relates to the utility scores assigned to prostate cancer health states in the model. Several of these were based on meta-analyses of studies conducted in different countries and settings. This is against best practice as the values cannot be considered to be equivalent when measured in different populations.[202] Some of the utility values may also be considered outdated or not relevant to a UK population.

5.3.3. Recommendations for future research

Although the calibrated model has predicted outcomes that are close to observed data, future research could explore the calibration space in greater detail to ascertain whether a better fit to data from the CAP trial could be achieved. Aspects of the model that were left unchanged from previous versions include the clinical detection rates; the average time from onset to metastatic cancer and from prostate cancer diagnosis to death; management of negative biopsies; and survival by stage, Gleason score, PSA values and age. Changes to these parameters were deemed unnecessary following the fit achieved by initial model adaptations but further work could explore the impact of adjusting these parameters to match with UK data.

This analysis has used the most recently available data from the ONS and the CAP trial as calibration targets however the calibration should be updated as new data become available. Data from the 15-year follow up of the CAP trial will provide greater insight into the longer-term effect of screening.

Clinical practice and diagnostic pathways for men at risk of prostate cancer continue to evolve, including the use of trans-perineal biopsy with image registration [329] and radio-labelled prostate-specific membrane antigen PET CT.[330 331] Future research could adapt the model to take account of newer innovations as they arise.

Similar to the work of Callender and colleagues[28 293], this analysis assumed that a higher risk score impacts only on the risk of developing cancer, with no impact on progression, and that genetic samples for all men would be available at low cost of acquisition (£25). In terms of risk thresholds, this analysis has taken the threshold found to be optimal by Callender et al[28] as a starting point. In reality, the optimal risk threshold, and that which would be deemed acceptable by men and their caregivers in terms of the trade-off between overdiagnosis and prostate cancer deaths averted, is still uncertain. Additionally, this analysis assumed that with risk-stratified screening there will be no screening in lower risk groups while a common screening strategy for those at higher risk will be adopted. A more appropriate screening strategy might involve different strategies for those above and below the 10-year prostate cancer risk threshold e.g., 2-yearly screening in those below the risk threshold and quadrennial screening in those above. This analysis, unlike that of Hendrix et al[292], did not explore the option of different strategies for different risk cohorts. Further research is needed in the area of polygenic risk scores and risk-stratified screening to explore different strategies and more accurately model these options.

5.4. Conclusions

This chapter details how a natural history model of prostate cancer has been adapted for use in a UK setting, thus enabling the comparison of the impact of novel screening strategies on the lifetime effectiveness and cost-effectiveness of prostate cancer screening. The

calibrated natural history model has shown good prediction of prostate cancer incidence in the UK and recent UK data have been identified to ensure the model is representative of a UK population. The work has drawn from the systematic review of recent cost-effectiveness models in prostate cancer screening to identify the best available natural history model and sources of data for model parameters. It has also drawn from the modified-Delphi consensus process which indicated the relevance of risk-stratified screening.

CHAPTER 6. COST-EFFECTIVENESS ANALYSIS

6.1. Introduction

This chapter draws on the previous chapters to analyse the cost-effectiveness of the screening strategies identified in the consensus process of Chapter 4, using the adapted and calibrated model of Chapter 5. The long-term costs, health outcomes and cost-effectiveness of prostate cancer screening in the UK is explored to determine the potential for screening to be effective and cost-effective compared to the current approach of no organised screening.

Cost-effectiveness analyses such as these are essential to make rational decisions about the allocation of limited healthcare resources. However, to date, robust evidence on the long-term cost-effectiveness of recent developments in prostate cancer screening alone, or in combination with one another at a national level within a screening programme, is lacking. The results of this analysis should assist healthcare policy makers to make informed decisions regarding the use of new prostate cancer screening innovations in a UK national screening strategy.

Any cost-effectiveness analysis comes with the caveat of uncertainty, particularly in the case of screening interventions. Policy makers should also be aware of the extent of this uncertainty when making decisions. This chapter demonstrates the use of established methods of dealing with parameter uncertainty in cost-effectiveness models, including deterministic and probabilistic sensitivity analysis.[332] The results provide an overview of how confident a policy maker could be in making a decision based on these results and where the key areas of uncertainty lie. Parts of this work have been published as a manuscript in Pharmacoeconomics.[333] This chapter expands on the methods and results described in the manuscript.

6.2. Methods

The analysis uses the adapted and calibrated natural history model described in Chapter 5 to estimate the lifetime cost-effectiveness, in terms of cost per QALYs gained, of prostate cancer screening from the perspective of the UK NHS. This perspective was chosen as it is in line with the NICE reference case.[89] The model uses 10-year data from the CAP randomised controlled trial to inform the natural history of the disease, as well as literature based sources for other parameters such as the costs and QALYs associated with prostate cancer testing and treatment (as detailed in Chapter 5).

6.2.1. Strategies compared

The strategies compared in the base case cost-effectiveness analysis reflected those identified as relevant by the panel of experts in the modified Delphi consensus process (Chapter 4).[242] These are shown in *Table 16*. As consensus was not reached on exact age ranges to screen or specific screening intervals, only that age-based screening should be considered, a range of age-based screening strategies based on those commonly compared in previous cost-effectiveness models were included.[231 334 335] The starting and stopping ages reflected those used in the CAP trial.[23] As the group of experts indicated that risk-based screening strategies should be compared, two strategies were included where screening starting age was based on polygenic risk score. An adaptive strategy was also included, based on the expert consensus, where screening interval was based on PSA score.

Table 16. Screening strategies compared in base case

Title	Starting age (years)	Stopping age (years)	Repeat screening interval	Comment
No screening	NA	NA	NA	
Screen 50	50	NA	None. Once-off screen.	
Screen 60	60	NA	None. Once-off screen.	
Screen 70	70	NA	None. Once-off screen.	
Repeat screen every 4 years	50	70	4-yearly	
Repeat screen every 2 years	50	70	2-yearly	
Risk-stratified 4-yearly	Age at which 10-year risk of developing prostate cancer is 7.5%, based on polygenic risk score	70	4-yearly	Based on most cost-effective strategy identified in recent analysis by Callender et al.[28]
Risk-stratified 2-yearly	Age at which 10-year risk of developing prostate cancer is 7.5%, based on polygenic risk score	70	2-yearly	To compare to age-based 2-yearly screening strategy.
Adaptive screening	50	70	PSA level of < 1.5 ng/ml screened every 6 years, with value > 1.5 ng/ml resulting in four-yearly screening	Based on ProScreen trial[336]

Scenario analyses were also conducted to observe the impact of using different ages to start and stop screening, screening intervals, and risk thresholds. The scenario analyses tested, with their rationale, are detailed in *Table 17*.

Table 17. Scenario analyses tested

Title	Starting age (years)	Stopping age (years)	Repeat screening interval	Comment
Repeat screen every 4 years 55-70 vs 50-70	55	70	4-yearly	To match screening starting age used in PLCO trial[22]
Repeat screen every 2 years 55-70 vs 50-70	55	70	2-yearly	
Adaptive screening 55-70 vs 50-70	55	70	PSA level of < 1.5 ng/ml screened every 6 years, with value > 1.5 ng/ml resulting in four-yearly screening	
Repeat screen every 4 years 50-74 vs 50-70	50	74	4-yearly	To match screening stopping age used in PLCO[22] and ERSPC[21] trials
Repeat screen every 2 years 50-74 vs 50-70	50	74	2-yearly	
Adaptive screening 50-74 vs 50-70	50	74	PSA level of < 1.5 ng/ml screened every 6 years, with value > 1.5 ng/ml resulting in four-yearly screening	
Repeat screen every 3 years	50	70	3-yearly	To capture all screening intervals between 2 and 5 years
Repeat screen every 5 years	50	70	5-yearly	
Risk-stratified 4-yearly 5% risk threshold vs 7.5%	Age at which 10-year risk of developing prostate cancer is 5%, based on polygenic risk score	70	4-yearly	To explore variation from 7.5% risk threshold
Risk-stratified 4-yearly 10% risk threshold vs 7.5%	Age at which 10-year risk of developing prostate cancer is 10%, based on polygenic risk score	70	4-yearly	
Risk-stratified 2-yearly 5% risk threshold vs 7.5%	Age at which 10-year risk of developing prostate cancer is 5%, based on polygenic risk score	70	2-yearly	
Risk-stratified 2-yearly 10% risk threshold vs 7.5%	Age at which 10-year risk of developing prostate cancer is 10%, based on polygenic risk score	70	2-yearly	
Adaptive screening intervals of 4 years and 2 years vs 6 years and 4 years	50	70	PSA level of < 1.5 ng/ml screened every 4 years, with value > 1.5 ng/ml resulting in two-yearly screening	To explore variation in intervals

In the no screening strategy, no organized or opportunistic testing was assumed. In all other strategies, men with a PSA value ≥ 3 ng/ml received a pre-biopsy mpMRI and combined

systematic biopsy and MRI-targeted biopsy if a Prostate Imaging–Reporting and Data System (PI-RADS)[337] value of 3-5 was found. This is in line with the results of the consensus process that MRI should be considered in screening pathways and also with recent guidance from NICE that MRI should be offered as the first-line investigation for people with suspected clinically localised prostate cancer.[157] For each symptomatic diagnosis, no screening focused PSA-testing but an average of two diagnostic 10-12 core transrectal ultrasound (TRUS)-guided biopsies was assumed. This is in accordance with evidence to suggest that repeat biopsies are common in men with an initial negative biopsy.[338-340]

6.2.2. Model

As detailed in Chapter 5, the model simulates a cohort of men to track prostate cancer onset and progression over time and then applies a screening strategy to this cohort to determine the change in outcomes (Figure 13). For this chapter, the life histories of 10 million men born in 1950 were simulated. As data from the CAP trial were used to calibrate the model, the simulated cohort reached age 50 in the year 2000 to correspond with the beginning of the CAP trial.

6.2.3. Outcome measures

Utility values were assigned to biopsy, a diagnosis of cancer, treatment with prostatectomy or radiation therapy, active surveillance, the post-recovery period, palliative therapy and terminal illness. In the absence of a single comprehensive data source from which to inform the health state values, the values used in the base case were sourced from a systematic review carried out by Hao et al[297], as described in Chapter 5. The values used are based on estimates measured using the EQ-5D instrument collected from various participant groups and in different settings. The limitations of this are discussed in Chapter 5.

6.2.4. Cost and resource use

Resource use related to PSA testing, polygenic risk stratification, biopsy, MRI, assessing suspected prostate cancer, treatment (prostatectomy or radiation therapy), active surveillance, palliative care and terminal illness. Costs are from the perspective of the UK

National Health Service (NHS) and correspond to the 2020 price year. Unit costs were based on recent UK-based sources including the NICE prostate cancer guideline[157] and a cost-effectiveness analysis by Callender et al (Chapter 5, *Table 14*).[28]

6.2.5. Analysis

The outcomes reported include the estimated number of pre-biopsy mpMRIs, prostate biopsies, prostate cancer incidence and prostate cancer deaths per 10,000 men for each strategy. The results are reported from age 30 over a lifetime horizon. Costs and QALYs for each strategy are reported and plotted on the cost-effectiveness plane with QALYs on the x-axis and costs on the y-axis. A cost-effectiveness frontier connects points on the plane to indicate which strategy (or pair of strategies) has the lowest cost per QALYs gained; strategies above and to the left of the frontier should be rejected on the grounds that they are not cost-effective. A steep gradient between successive points on the frontier indicates a high cost per QALY gained for the more costly strategy.

ICERs, which are calculated by dividing incremental costs by incremental QALYs (defined in section **Error! Reference source not found.**), are reported for all interventions compared with the next non-dominated intervention. An intervention is considered to be dominated if it provides fewer mean QALYs at a higher mean cost than another intervention, and extendedly dominated if it provides fewer mean QALYs at a higher mean cost than a weighted average of 2 alternative interventions. An ICER below £30,000 per QALY gained was considered cost-effective in accordance with the NICE reference case.[89]

Net monetary benefits at thresholds of £20,000 and £30,000 per QALYs gained are also reported. When comparing net monetary benefit between alternative strategies, a higher estimate indicates that the strategy is cost-effective compared with its alternative, at the given willingness-to-pay threshold. A discount rate of 3.5% was applied to all future costs and QALYs, reflecting NICE guidance.[89]

Deterministic Sensitivity analyses

Deterministic sensitivity analyses were carried out to test the robustness of results to changes in key model parameters. The source used to inform the health state utility values, Hao et al[295], also reported utilities based on a systematic review and meta-analysis of studies using the prostate-cancer-specific Patient Oriented Prostate Utility Scale (PORPUS-U).[297] The EQ-5D health state values were used in the base case as they have the advantage of being general to any disease area and to correspond with the QALY norms used.[295] EQ-5D-5L is also the measure recommended by decision makers to facilitate comparison across disease areas and interventions.[89] However, it has been shown that disease-specific instruments can be more sensitive at detecting changes in patients with prostate cancer.[341] The PORPUS-U has 10 items (pain, energy, social support, communication with doctor, emotional well-being, urinary frequency, urinary leakage, sexual function, sexual interest and bowel function), each with four to six levels of severity.[342] Utility weights derived from prostate cancer patients are used to generate utility values (0=dead and 1=full health).[343] A sensitivity analysis using PORPUS-U estimates in place of EQ-5D estimates, where available, was therefore carried out. The estimates used and their data sources and method of estimation are shown in *Table 18*.

Table 18. PORPUS-U utility estimates and their sources

Parameter	PORPUS-U estimate	95% CI	Source	Method	Participants
Prostatectomy part 1 (first 2 months)	0.86	(0.76 – 0.96)	Magnus 2019[312] (meta analysis of Krahn 2009[183] and Ku 2009[344])	Krahn 2009: PORPUS-U collected before prostatectomy and 2 and 12 months post treatment. Ku 2009: PORPUS-U collected prior to prostatectomy, and at 0 to 3 months, 3 to 9 months, 9 to 18 months, and 18 to 30 months post-treatment	Krahn 2009: 134 Canadian patients diagnosed with prostate cancer within previous 6 months and scheduled to receive prostatectomy Ku 2009: 213 Canadian patients with clinically localized prostate cancer
Prostatectomy part 2 (next 10 months)	0.90	(0.84 - 0.97)	Magnus 2019[312] (meta analysis of Krahn 2009[183] and Ku 2009[344])	Krahn 2009: PORPUS-U collected before prostatectomy and 2 and 12 months post treatment. Ku 2009: PORPUS-U collected prior to prostatectomy, and at 0 to 3 months, 3 to 9 months, 9 to 18 months, and 18 to 30 months..	Krahn 2009: 68 Canadian patients diagnosed with prostate cancer within previous 6 months and scheduled to receive prostatectomy Ku 2009: 213 Canadian patients with clinically localized prostate cancer
Radiation therapy part 1 (first 2 months)	0.89	(0.87 - 0.91)	Krahn 2009[183]	PORPUS-U collected before radiation therapy and 2 and 12 months post treatment.	66 Canadian patients diagnosed with prostate cancer within previous 6 months and scheduled to receive radiation therapy
Radiation therapy part 2 (next 10 months)	0.92	(0.90 - 0.94)	Krahn 2009[183]	Described above	Described above
Active surveillance	0.98	(0.97 - 0.99)	Loeb 2018[307]	PORPUS-U completed prior to focus group discussion	36 men with prostate cancer on active surveillance from 2 US hospitals
Post recovery period	0.93	(0.91 - 0.95)	Magnus 2019[312] (meta-analysis of Avila 2014[345], Bremner 2014[346], Krahn 2013[347], Ku 2009[344])	Avila 2014: PORPUS-U collected annually via telephone Bremner 2014: PORPUS-U collected via post Krahn 2013: PORPUS-U collected via post	Avila 2014: 480 Spanish men diagnosed with prostate cancer up to 10 years previously. Bremner 2014: 676 Canadian men diagnosed with prostate cancer up to 10 years previously

				Ku 2009: PORPUS-U collected at 0 to 3 months, 3 to 9 months, 9 to 18 months, and 18 to 30 months post-treatment	Krahn 2013:585 Canadian men diagnosed with prostate cancer up to 10 years previously Ku 2009: 213 Canadian patients with clinically localized prostate cancer
Palliative therapy	0.68	(0.64 – 0.71)	Magnus 2019[312] (Meta-analysis of Farkkila 2014[319] and Torvinen 2013[187], 15D values)	Farkkila 2014: 15D instrument Finnish with valuation algorithm for utility scores. Torvinen 2013: 15D instrument Finnish with valuation algorithm for utility scores.	Farkkila 2014: 30 prostate cancer patients in Helsinki with metastatic disease and receiving palliative treatments only or who died due to cancer within 6 months of responding to the questionnaire Torvinen 2013: 19 men with prostate cancer from the Helsinki and Uusimaa Hospital District receiving palliative care
Terminal illness	Used EQ-5D estimate				
Biopsy	Used EQ-5D estimate				
Cancer diagnosis	Used EQ-5D estimate				

One-way sensitivity analyses were also carried out on cost and utility parameters including the unit costs and disutilities associated with prostate biopsy, mpMRI and treatment. These were varied by 20% or to an upper bound of 1 in the case of utilities.

Probabilistic analysis

Probabilistic analysis is a method of dealing with uncertainty in model inputs and quantifying the level of confidence in model results, in relation to this uncertainty, by sampling parameters from their respective distributions (rather than simply using the mean parameter values). A probabilistic analysis was carried out using 1000 replicates for a population of one million men to address the impact of uncertainties in the test accuracies,

costs, health state values and natural history parameters. The uncertainty distributions assumed are given in Chapter 5. Test accuracies and health state values were assumed to be normally distributed on the logit scale. Test characteristics were assumed to be independent. Costs were sampled from a gamma distribution with mean 1 and 95% confidence interval between 0.8 and 1.22. The natural history parameters were assumed to follow a multivariate normal distribution.

Results are presented using cost-effectiveness acceptability curves, which reflect the probability of a strategy being most cost-effective at each willingness to pay per QALYs gained threshold. The probability of each strategy being optimal at £20,000 and £30,000 per QALY gained was calculated by counting the proportion of samples for which the expected net benefit was highest. The incremental costs and QALYs compared to no screening estimated for each strategy at each iteration are also plotted on the cost-effectiveness plane.

The sampled parameters were also used to explore how uncertainty in the model inputs impacts on the intervention considered to be optimal using value of information (VoI) methods.[348] The expected value of perfect information (EVPI) gives an upper bound on the benefit in reducing uncertainty in all of the model inputs while the expected value of partially perfect information (EVPPI) gives an upper bound on the benefit in reducing uncertainty in a subset of the inputs, highlighting the parameters to which the decision is most sensitive. EVPI and EVPPI were computed per person for willingness to-pay per QALY thresholds of £20,000 and £30,000 and multiplied by the estimated annual incidence of prostate cancer in the UK of 52,000[16] to obtain population-level EVPI and EVPPI. The population EVPI over 1-year, 10-year and 15-year time horizons was calculated. The Sheffield Accelerated Value of Information web application[349] was used to compute EVPPI for subsets of parameters.[350]

Finally, Expected Loss Curves (ELCs) were plotted which present the consequences of choosing a suboptimal strategy in terms of expected foregone benefits, plotted as a function of willingness to pay per QALYs gained.[351] ELCs display the optimal strategy, the value of

eliminating decision uncertainty through additional research, and the ranking of strategies in terms of expected losses.[230]

6.3. Results

6.3.1. Health outcomes

Table 19 presents predicted outcomes per 10,000 men simulated over a lifetime from the age of 30. Compared with no screening, the screening strategies resulted in more biopsies, more prostate cancers diagnosed and fewer prostate cancer deaths. Overdiagnosis, defined as prostate cancer diagnoses in men who would never have been diagnosed without screening, was shown to increase with more intensive screening with a once-off screen at 50 resulting in only two overdiagnosed cases per 10,000 men but repeat screening every 2 years resulting in 113 overdiagnosed cases. Compared with 4-yearly age-based screening, risk-stratified screening with 4-yearly intervals was associated with 46% fewer overdiagnosed cancers, 66% fewer pre-biopsy mpMRIs and 63% fewer screen-initiated biopsies. The reduced costs and harms were at the expense of 8% more prostate cancer deaths.

Table 19 Predicted outcomes and health economic results from age 30 by strategy

	No screening	Screen 50	Screen 70	Screen 60	Risk-stratified 4-yearly to age 70	Risk-stratified 2-yearly to age 70	Adaptive screening 50-70	Repeat screen every 4 years 50-70	Repeat screen every 2 years 50-70
Outcomes per 10,000 men*									
Number of screens	0	8285	7029	7978	6173	10049	31155	40223	65335
Number of MRI events	0	96	1022	627	834	1205	2458	2423	3519
Number of screen-initiated biopsies	0	64	631	400	538	750	1471	1452	2013
Number of clinically initiated biopsies	2817	2752	2474	2491	2390	2295	1925	1942	1826
Diagnosed PCa	1487	1488	1547	1509	1538	1555	1583	1582	1600
Screen diagnosis	0	36	245	195	278	345	569	558	636
Metastatic cancer	254	248	217	226	215	207	173	174	167
Localised & G<6	702	705	757	731	757	774	805	803	823
Localised & G=7	292	296	313	308	313	318	336	336	339
Localised & G>7	238	239	260	244	252	256	268	269	271
Overdiagnosis	0	2	60	22	51	68	97	95	113
PCa death	527	520	512	504	497	489	457	459	444
Life years, QALYs, Costs, ICERs and NMB per man. Strategies sorted by increasing costs*									
Life years (undiscounted)	51.5370	51.5506	51.5515	51.5671	51.5707	51.5785	51.6237	51.6212	51.6388
Life years, 3.5% discounted	23.5624	23.5652	23.5646	23.5676	23.5680	23.5692	23.5775	23.5771	23.5801
QALYs (undiscounted)	42.518	42.527	42.519	42.534	42.532	42.535	42.560	42.559	42.567
QALYs, 3.5% discounted	20.2428	20.2442	20.2419	20.2440	20.2431	20.2429	20.2452	20.2450	20.2451
NHS costs, undiscounted	£2,013	£2,042	£2,152	£2,117	£2,161	£2,224	£2,414	£2,424	£2,610
NHS costs, 3.5% discounted	£449	£466	£488	£493	£506	£528	£609	£613	£685
ICER		£12,860	Dominated	Extendedly dominated	Extendedly dominated	Extendedly dominated	£137,364	Dominated	Extendedly dominated
Net Monetary Benefit (£20,000/QALY)[∞]	£404,408	**£404,417**	£404,351	£404,386	£404,356	£404,329	£404,296	£404,286	£404,£217
Net Monetary Benefit (£30,000/QALY) [∞]	£606,836	**£606,859**	£606,771	£606,826	£606,787	£606,758	£606,748	£606,735	£606,668

**Based on life histories of 10 million simulated men[∞]Strategy with highest net monetary benefit highlighted in bold*

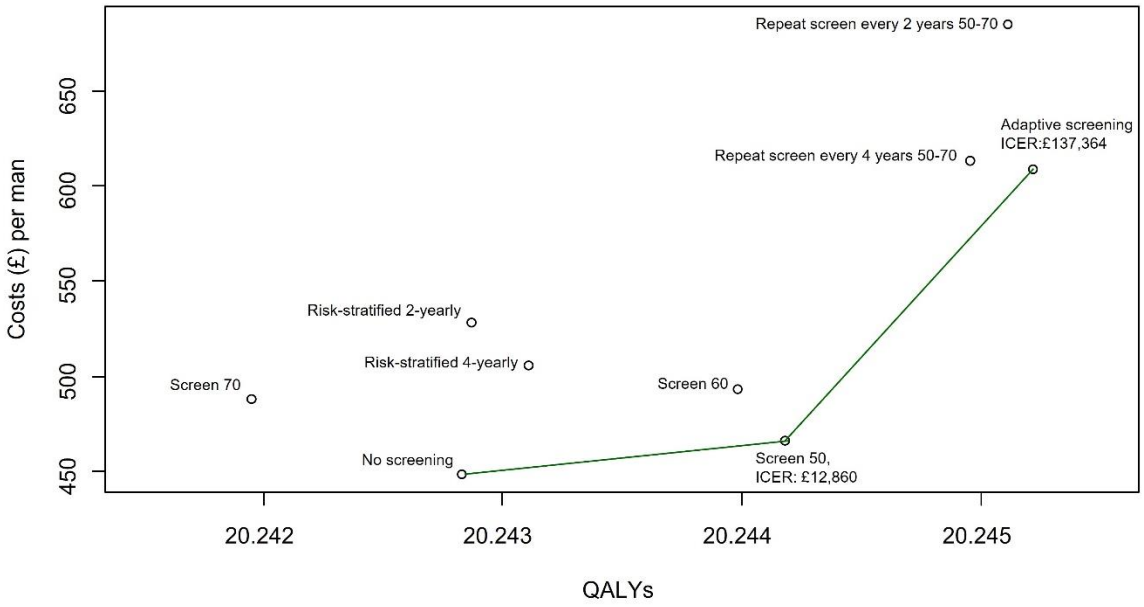
Abbreviations. MRI=magnetic resonance imaging. PCa=prostate cancer. G=Gleason score. QALYs=quality adjusted life years. NMB=net monetary benefit. NHS=National Health Service, ICER: Incremental Cost-Effectiveness Ratio, PSA: Prostate Specific Antigen.

6.3.2. Costs and QALYs

The strategies in Table 19 are sorted by predicted mean costs. The lowest costs were found for no PSA screening and the highest for repeat PSA screening every 2 years for all men between the ages of 50 and 70 years. All screening strategies other than a once-off screen at 70 resulted in a slight increase in QALYs (range: 0.0001 – 0.0024) compared to no screening. Compared to no screening, the mean net monetary benefit was lower for all screening strategies at both a £20,000 and £30,000 per QALYs gained threshold, other than a once-off screen at 50. Note that the risk-stratified strategies, although not cost-effective compared to no screening, had higher net monetary benefits than their age-based equivalents (e.g. risk-stratified 4-yearly screening to age 70 had a higher net monetary benefit than repeat screening every 4 years to age 70).

The strategies on the cost-effectiveness frontier, as shown in Figure 20, were no screening, a once-off screen at 50 and adaptive screening. The ICER of moving from a policy of no screening to a once-off screen at 50 is £12,860 per QALY gained, which is under the £20,000-£30,000 per QALY gained threshold recommended by NICE and therefore considered to be cost-effective. Adaptive screening, although on the frontier, would not be considered cost-effective due to the relatively high ICER of moving to this strategy from a once-off screen at 50 (£137,364 per QALY gained).

Figure 20. Cost-effectiveness frontier

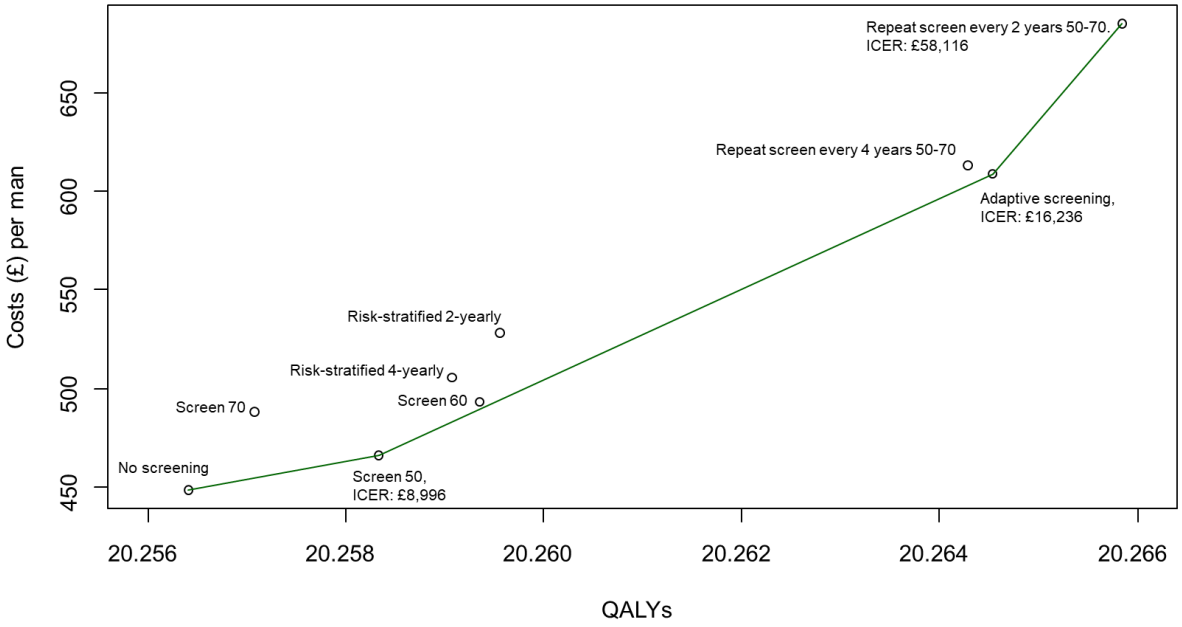


6.3.3. Sensitivity analyses

One-way and scenario analyses

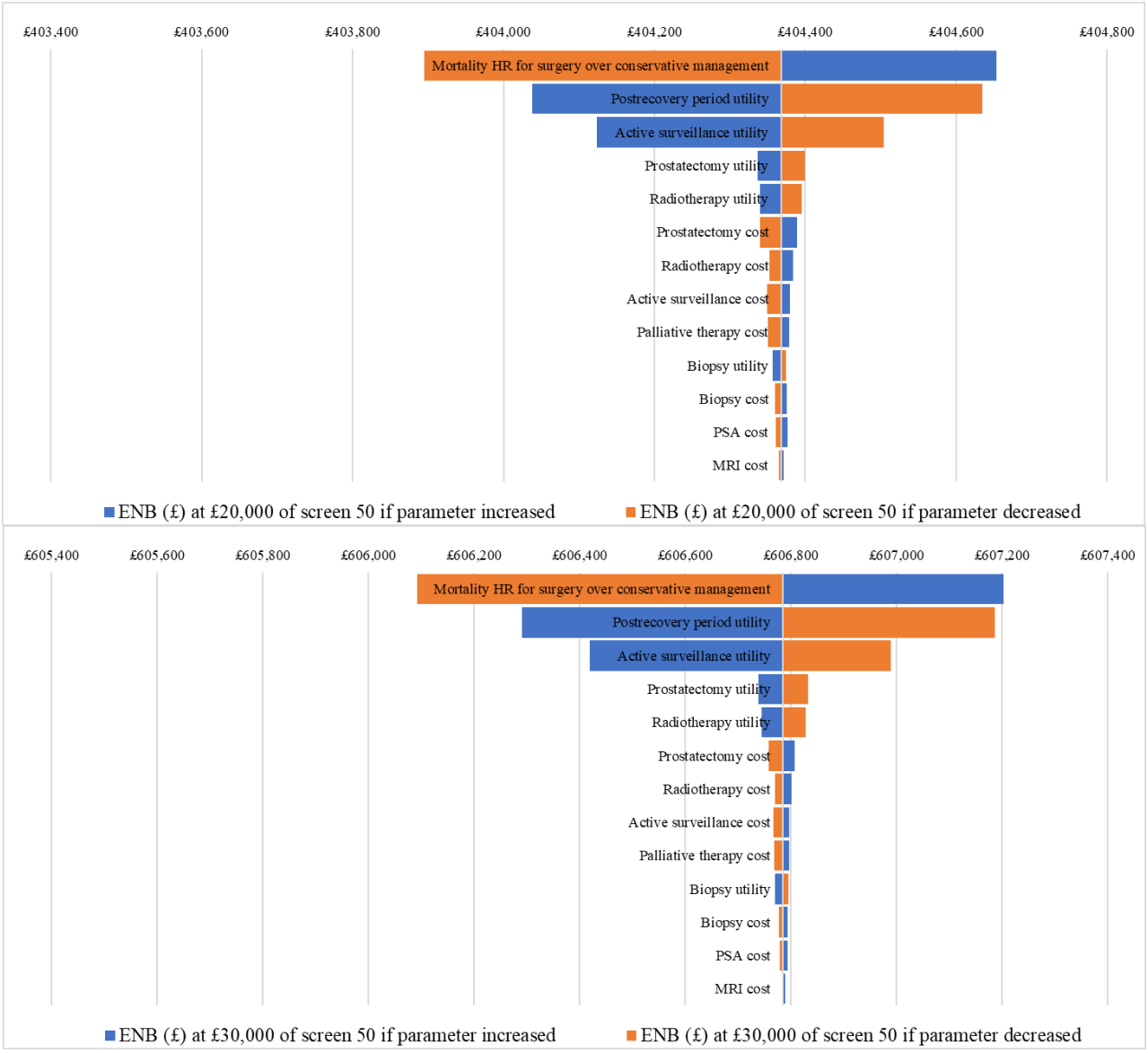
Figure 21 plots the results of the sensitivity analysis using the disease specific (PORPUS-U) utility estimates rather than the EQ-5D estimates on the cost-effectiveness frontier. Applying the PORPUS-U weights leads to a change in the optimal policy choice. A once-off screen at 50 years old is still found to be cost-effective compared to no screening with an ICER of £8,996 per QALY gained. However, in this sensitivity analysis, the ICER for the comparison of adaptive screening to a once-off screen at 50 is lower than the £20,000 – £30,000 willingness to pay threshold recommended by NICE at £16,236 per QALYs gained and would therefore be recommended. Repeat screening every 2 years from 50-70 years old is also on the cost-effectiveness frontier with an ICER compared to adaptive screening of £58,116 per QALY gained.

Figure 21. Cost-effectiveness frontier assuming utility estimates measured with PORPUS-U



The one-way sensitivity analyses further highlight the importance of the utility estimates used. The impact on the ENB of a once-off screen at 50 at the £20,000 and £30,000 willingness to pay thresholds is shown in Figure 22. Of the parameters tested, those with the biggest impact were the mortality hazard ratio for surgery over active surveillance and the utilities associated with the post-recovery period and active surveillance. The utility decrements associated with these health states are influential in the model as they are assumed to last for 7 and 9 years respectively. Altering the costs and utilities associated with treatment and diagnosis had a negligible impact on the ICER.

Figure 22. One-way sensitivity analyses comparing impact of varying model parameters on ENB of a once-off screen at 50



**Costs and utilities varied by 20% or to an upper bound of 1 in the case of utilities. Mortality hazard ratio for surgery over conservative management varied to upper and lower bounds of estimate in ProtecT trial (Mean 0.63, 95% CI: 0.21 - 1.93)*

Regarding the scenario analysis considering different start and stop ages, intervals and risk thresholds, Figure 23 shows the expected net benefit of the alternative strategies tested compared to their respective base case scenarios. Most alternative strategies resulted in negative expected net benefit, which means they would be considered less cost-effective. The exceptions were the scenario analyses considering risk thresholds of 5% and 10% as opposed to 7.5% in the risk-stratified screening strategies. At a £30,000 willingness to pay per QALYs gained threshold, the strategies using risk thresholds of 5% resulted in increased

net benefit compared to those using a threshold of 7.5%. Conversely, at a £20,000 per QALYs gained willingness to pay threshold, the strategies using risk thresholds of 10% resulted in increased net benefit relative to those using a 7.5% threshold. This suggests that, as willingness to pay per QALYs gained increases, the most cost-effective strategies are those which have a lower risk threshold for screening, resulting in an increasing number of men being screened. These results suggest that further research around optimal risk thresholds is warranted to define the exact risk threshold that may be most cost-effective.

Figure 23. Variation in Expected Net Benefit for each scenario analysis vs respective base case scenario

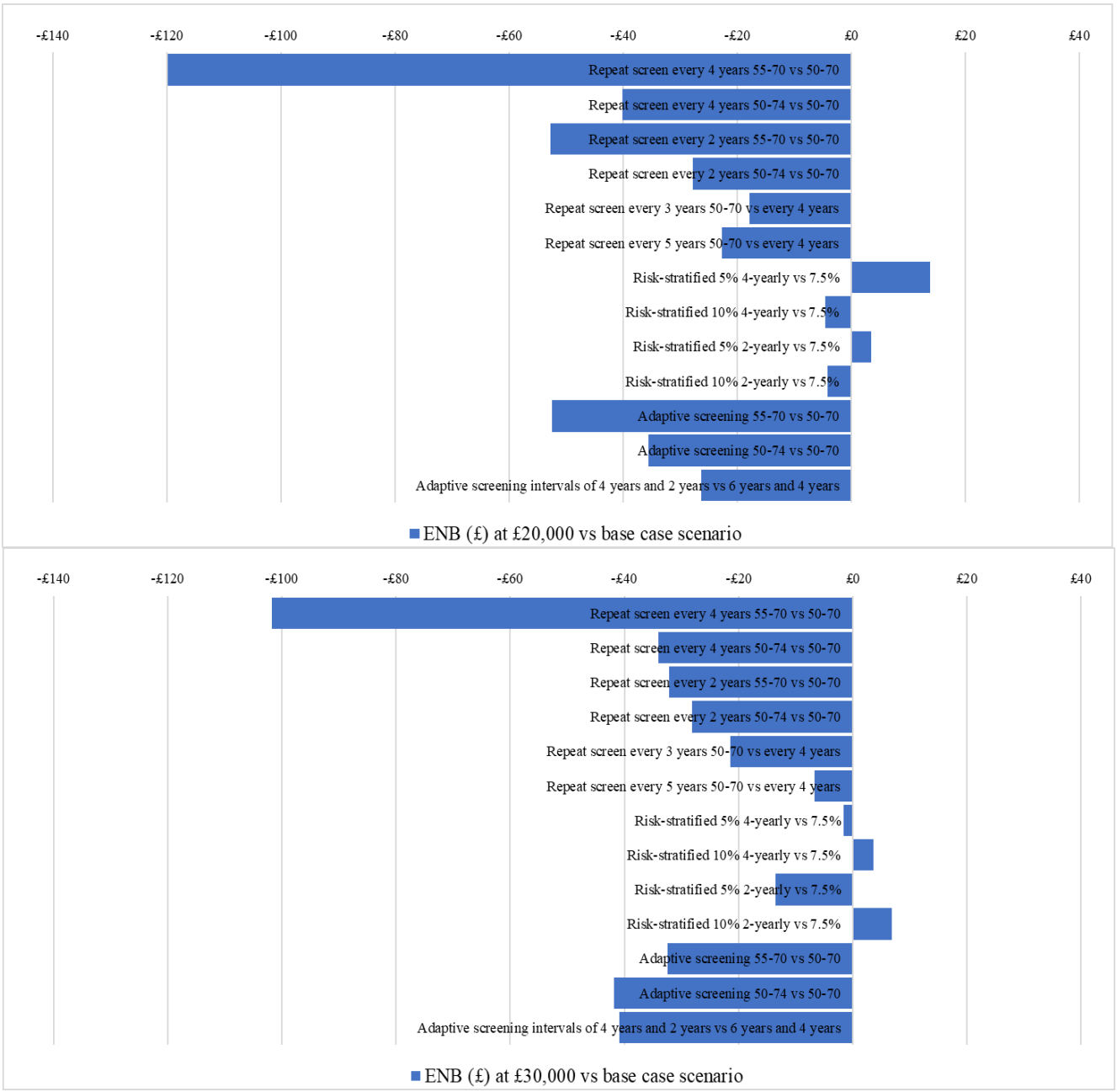


Figure 24 demonstrates the impact of the alternative strategies on life years gained. As expected, more intensive strategies result in increased life years.

Figure 24. Variation in life years gained for each scenario analysis vs respective base case scenario

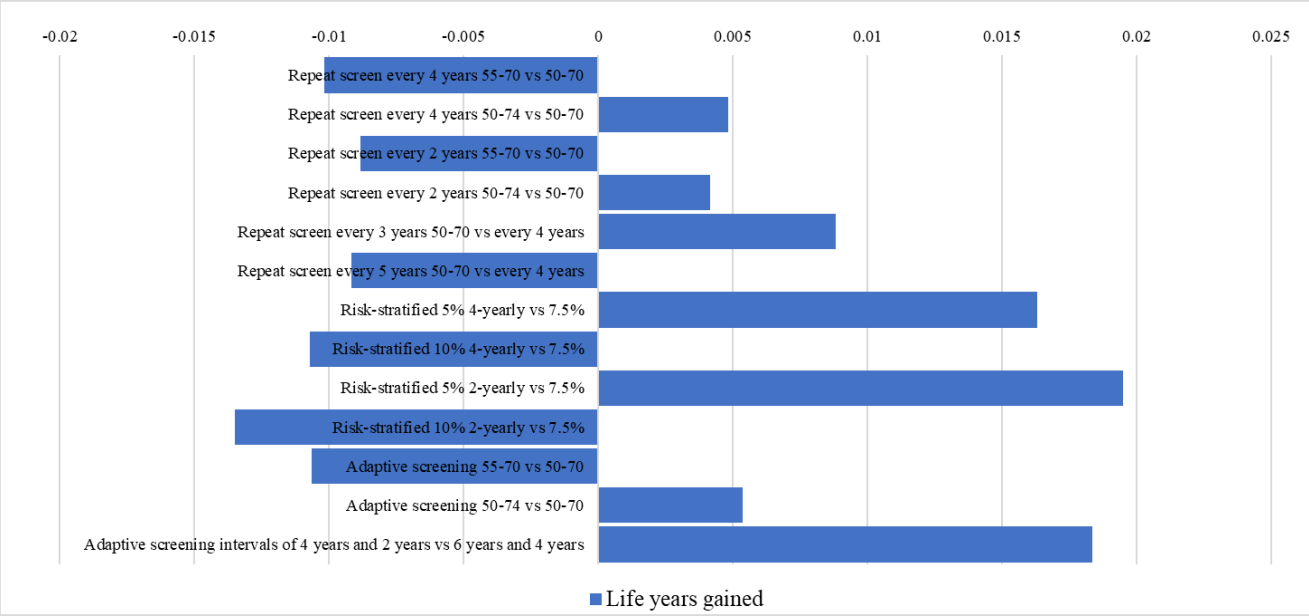
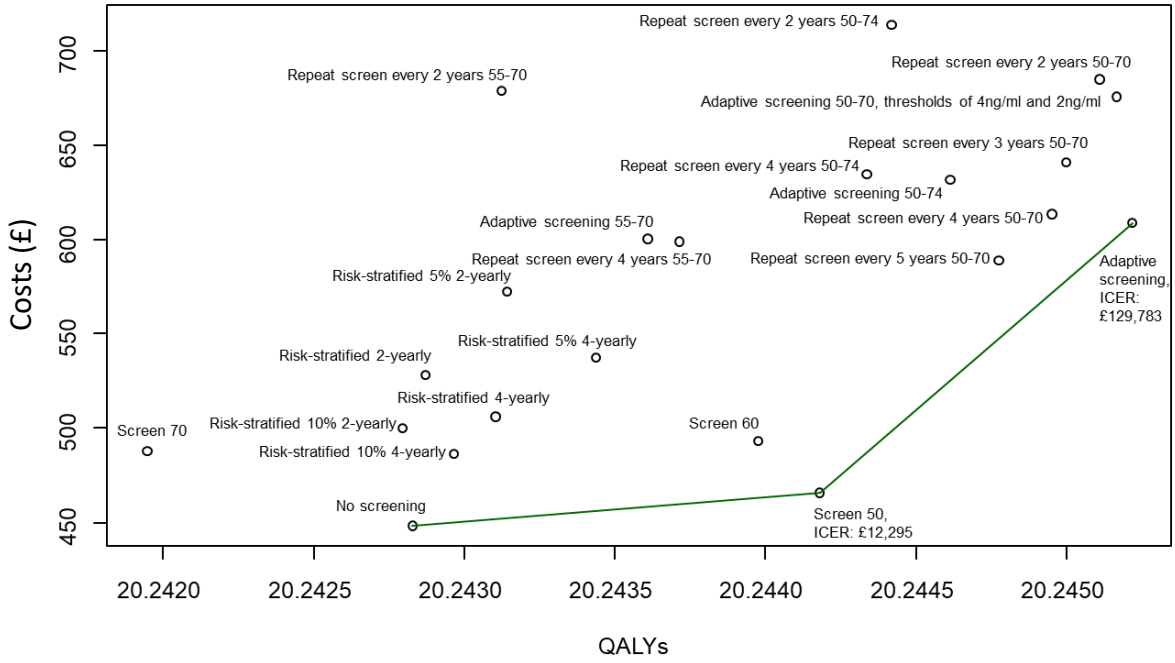


Figure 25 plots the predicted costs and QALYs from all strategies tested on the cost-effectiveness frontier. This shows that the strategies on the frontier are unchanged from those considered in the base case.

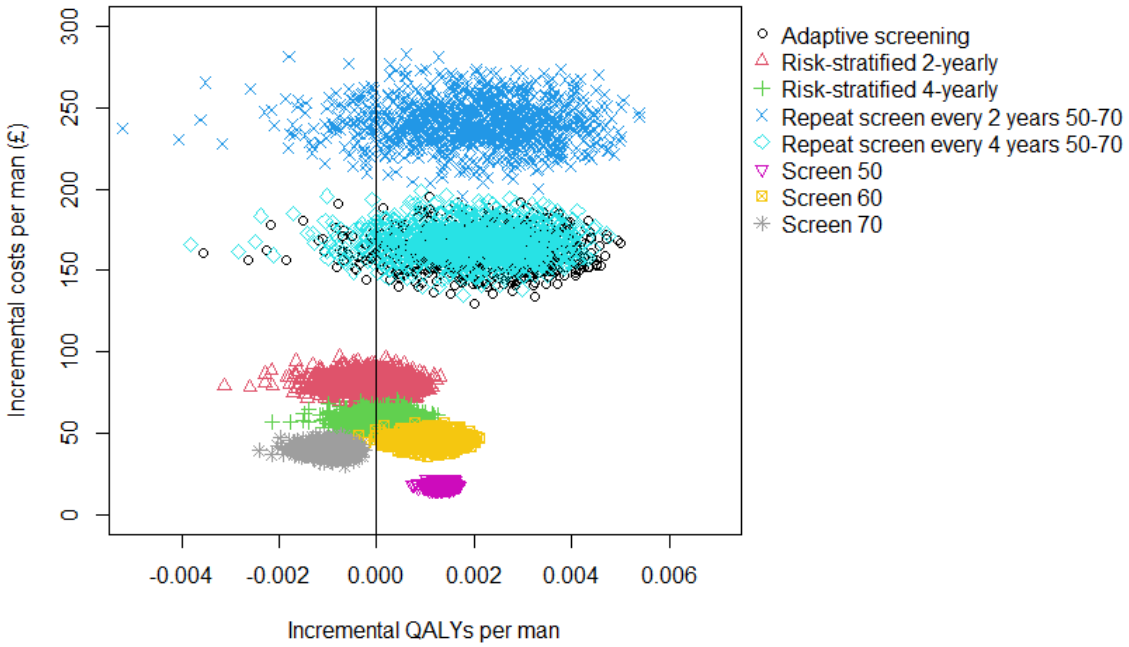
Figure 25. Cost-effectiveness frontier with scenario analyses included



Probabilistic analysis

Figure 26 plots incremental costs and effects at each probabilistic iteration for all strategies compared to no screening on the cost-effectiveness plane. This shows the relatively small estimated costs of a once-off screen at 50 compared to the other strategies. This is also the only strategy not to have the possibility of negative incremental QALYs gained compared to no screening. Mean costs increase with increased screening frequency, as does the estimated uncertainty in costs. This reflects the increased numbers of prostate cancers diagnosed with increased screening and the associated uncertainty in costs, utilities and natural history parameters.

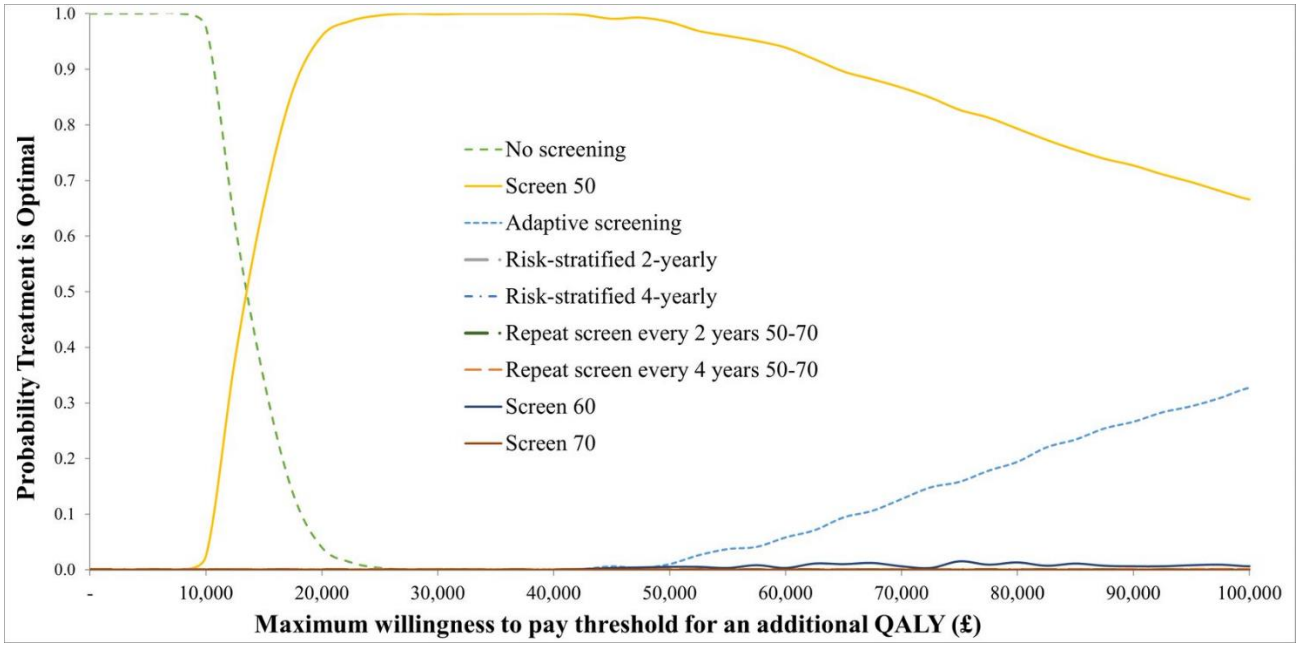
Figure 26. Cost-effectiveness plane plotting incremental costs and effects at each iteration for all strategies compared to no screening



The cost-effectiveness acceptability curves shown in Figure 27 highlight the separation of strategies by the willingness to pay threshold. At any willingness to pay threshold above £15,000 per QALY gained, the strategy with the highest probability of being optimal is a once-off screen at 50. At a willingness to pay threshold of £20,000 per QALY gained a once-off screen at 50 has a 95.5% probability of being optimal with no screening having a 4.5%

chance. At a threshold of £30,000 per QALY gained the probability that a once-off screen at 50 is optimal increases to 99.9%.

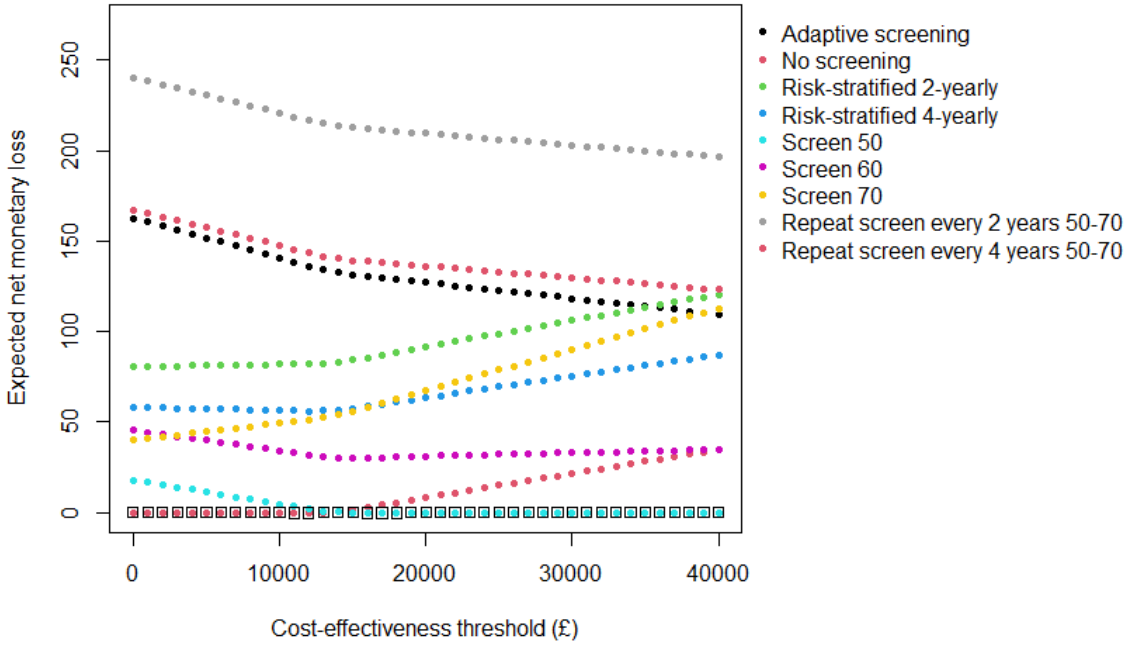
Figure 27. Cost-effectiveness acceptability curves



With regard to the VOI analysis, at a willingness to-pay per QALY threshold of £20,000 the per-person EVPI was only 1 pence and the population EVPI, representing all new prostate cancers diagnosed in the UK, was £629 over a 1-year time horizon. The population EVPI over 10-year and 15-year time horizons was £5415 and £7498 respectively. At a £30,000 threshold these values were £0. These values are small as they represent only the parameter uncertainty assumed in the probabilistic analysis and the decision is relatively clear cut at these thresholds, as shown in Figure 27. The one-way and scenario analyses have shown that the decision of which strategy to recommend is sensitive to more significant changes in parameter values, particularly utilities. The parameter with the largest EVPPI was the EQ-5D based utility associated with active surveillance, which was also highlighted in the one-way sensitivity analysis as influential.

The expected loss curves in Figure 28 echo the cost effectiveness acceptability curves by showing that expected net loss is lowest for no screening up to a willingness to pay threshold of around £15,000 per QALY, at which point the strategy with the lowest expected net loss is a once-off screen at 50.

Figure 28. Expected loss curves



6.4. Discussion

In summary, of the strategies compared, a once-off PSA screen at 50 years old was the only strategy which would be considered cost-effective at the cost-effectiveness threshold of £30,000 per QALYs gained recommended by NICE in the base case. Although this strategy was estimated to avert only 7 prostate cancer deaths per 10,000 men compared to no screening, it also resulted in only 2 overdiagnosed cancers per 10,000 men compared to 113 overdiagnosed cases per 10,000 men in a strategy where screening is repeated every 2 years. The low ICER of £12,860 suggests that this improvement in quantity and quality of life is enough to justify the increased costs of a once off screen at 50 (using PSA combined with pre-biopsy MRI) for all men at current UK willingness to pay per QALY gained thresholds. This is an important finding given the current recommendation against prostate cancer screening in the UK.

The finding was sensitive to the health state utility values used, with two sets of health state values available. In comparison to the estimates when using the disease-specific PORPUS-U measure, the EQ-5D estimates used in the base case assumed lower health related quality of life associated with prostatectomy, radiation therapy, active surveillance, the post recovery period and palliative therapy. This has the effect of making the diagnosis and treatment of prostate cancers less cost-effective. The sensitivity analysis using the PORPUS-U estimates showed that adaptive screening may be cost-effective if the health-related quality of life associated with these states is higher. In practice, the utility estimates based on the EQ-5D instrument are those that would be preferred by decision makers such as NICE, unless it could be shown that the EQ-5D does not perform as would be expected or is not responsive to changes in health in men with prostate cancer.[305] Nevertheless, it is important to note how a change in utility values could impact the strategy deemed most cost-effective.

6.4.1. Comparison with previous studies

Two previous UK cost-effectiveness analyses have found that a PSA screening strategy with risk-stratification by polygenic risk score was associated with an improvement in cost-

effectiveness compared with age-based screening.[28 293] The studies by Callender and colleagues compared age-based 4-yearly PSA screening to polygenic risk-based 4-yearly PSA screening at different risk thresholds with no comparison of once-off PSA screens or adaptive PSA screening. Making the same comparison in this analysis (polygenic risk-based 4 yearly vs. age based 4 yearly) gives similar results with the polygenic risk-stratified strategies having higher net monetary benefits than their age-based equivalents, although in the Callender studies the risk-stratified strategies are less costly and result in more QALYs than their age-based alternatives.

In both this research and that of Callender et al, risk-stratified screening was associated with fewer overdiagnoses and biopsies and lower costs than age-based screening but with more deaths from prostate cancer. The magnitude of expected differences between the strategies varied, with the most recent Callender et al paper finding risk-stratified PSA screening with the use of mpMRI-targeted biopsy to be associated with 60% fewer overdiagnosed cancers at a 7.5% risk threshold compared to 46% in this study.[28] Callender et al predicted around 50% fewer pre-biopsy mpMRIs and prostate biopsies compared with age-based screening, compared to around 65% in this analysis. They found 12% more deaths from prostate cancer with risk-stratified PSA screening compared to the finding from this work of 8%. These differences are likely due to the life-table based approach to modelling used in the Callender et al papers where, rather than modelling individual life histories and cancer progression, simplifying assumptions were made with regard to the effect of screening e.g. a 15% reduction in advanced cancer at diagnosis if screened. The model used by Callender et al also did not distinguish between Gleason grades or allow for progression between cancer stages.

A recent US cost-effectiveness analysis, Hendrix et al[292], comparing age-based screening to genetic risk-stratified screening concluded that risk-adapted screening was more cost-effective when compared to less intensive age-based screening. Risk-stratified strategies were cost-effective compared to biennial screening starting at age 55, for example, but not compared to biennial screening starting at age 45. This highlights the importance of comparing a wide range of screening strategies and intensities.

Two previous studies were identified comparing the cost-effectiveness of once-off and repeat universal screens to adaptive screening.[334 335] One study[334] assumed that men with PSA levels above the median for their age are rescreened in 2 years and the rest return in 4 years. That study also found that adaptive screening was on the cost-effective frontier when considering cost per life-year gained. Similar to the analysis carried out for this chapter, this result was sensitive to the utility estimates used. When utilities for locoregional disease were increased to their upper 95% confidence and utilities for distant disease and end-of-life states were decreased to their lower 95% confidence limit, all screening strategies resulted in a loss of QALYs. The other study[335] compared universal screening to adaptive strategies where the screening interval was every 1 year if the PSA level was higher than 3.0 ng/mL and every 2 years otherwise, or the screening interval was every 2 years if the PSA level was higher than 1.0 ng/mL and every 4 years otherwise. They found that the latter strategy compared favourably to no screening, with an ICER under the acceptable threshold, although the best performing strategies were non-adaptive.

Heijnsdijk et al used the MIcrosimulation SCreening Analysis (MISCAN) model to determine optimal prostate cancer screening intervals and ages based on data from ERSPC.[352] They found that screening strategies with intervals of three years are more cost-effective than those using longer intervals, whereas this study found an interval of 4 years to be preferable to 3 years. Both studies agreed that a lower age to stop screening is preferable with the scenario analysis for this study finding that stopping screening at 70 achieves a higher net benefit than stopping at 74.

The most recent UK cost-effectiveness analysis comparing once-off and repeat age-based PSA screening strategies identified was that of Hummel and Chilcott in 2013.[353] Hummel and Chilcott concluded that PSA screening was not effective compared with no screening. However, it is unclear whether their model was well calibrated to the 11-year prostate cancer mortality rate ratio from ERSPC.[354]

6.4.2. Strengths and limitations

The primary strengths of this analysis are: (1) the comparison of strategies in the base case that were chosen as relevant by the group of experts in the Delphi consensus process.[242] The strategies compared give a picture of the potential impact of relevant screening strategies in the UK today; (2) The comparison of strategies using a detailed natural history model which accounts for progression and has shown good prediction of prostate cancer incidence in the UK when compared to data from the CAP trial and the ONS; (3) the wide range of sensitivity and scenario analyses which have been explored to assess the robustness of results to parameter and decision question uncertainty.

The results of this analysis are particularly relevant given the current uncertainty around whether prostate cancer screening should be provided in the UK. [355] This is evidenced by the fact that the results were presented and discussed at a prostate cancer workshop organised by the National Screening Committee in June 2021 to discuss the future of prostate cancer screening in the UK, attended by over 30 experts and academics.

The results have shown that a once-off PSA screen at 50 years old has the potential to be cost-effective in a UK setting when compared to no screening. A limitation is that the analysis did not consider the substantial costs involved in setting up a screening programme, where none is currently in place. Such programme costs include costs incurred outside of the point of delivery of an intervention to patients such as implementation, organization, administration, monitoring, and evaluation.[356-358] It is common to exclude these in cost-effectiveness analyses due to costs being equivalent across comparators[357], however, in the case of moving from no screening to screening, programme costs could be substantial. Including these would likely reduce the potential for a once-off screen to be cost-effective.

A further limitation is an absence of an assessment of the reliability of the results to changes in model structure. The model health states categorise prostate cancer into Gleason grade ≤ 6, 7 or ≥7, for example, while research has shown that the distinction between 3+4 and 4+3 within Gleason grade 7 is important for prognosis and treatment allocation.[359] Due to data

limitations, the implications of changing the model structure to make this and other potential distinctions were not explored.

An additionallimitation is that the model assumes the use of pre-biopsy MRI in a screening setting, following NICE recommendations[157], and allocation of men to treatment with either radical prostatectomy, radiation therapy or active surveillance based on 2016 data from the National Cancer Registration and Analysis Service. The cost-effectiveness of alternative downstream diagnostic and treatment pathways is not explored. Future work could consider the cost-effectiveness of MRI in a UK setting, including assessing the various methods of administration identified in the systematic review (Chapter 3). Alternative treatments such as hormone therapy or brachytherapy, and differing allocation of treatments using more recent data, could also be explored.

6.4.3. Recommendations for future research

Although the base case analysis and scenario analyses have explored a wide range of screening strategies, there is scope for further exploration. As an absolute absence of PSA screening is unlikely to be achieved in countries where opportunistic screening is increasingly common, future research could consider comparing potential screening strategies to opportunistic screening, rather than no screening. This approach would require accurate data on current PSA testing, pre-biopsy MRIs and prostate biopsies. Future research might also explore the efficiency of other low-intensity strategies such as screening twice over a lifetime, considering the finding from this piece of work that a once-off screen at 50 is optimal. Additional exploration of adaptive strategies with different PSA thresholds and intervals may also be useful. Finally, although the comparison of screening strategies incorporating novel biomarkers was not identified as a priority in the consensus process (Chapter 4), there may be merit in assessing the cost-effectiveness of low-intensity screening strategies incorporating both a novel biomarker and MRI prior to biopsy.

The choice of screening model may affect the predicted cost-effectiveness. A stage-shift model, which was used in this analysis, assumes that the benefit associated with screening is due to a shift to a less advanced stage at diagnosis. Cure models assume that a proportion

of cancers detected earlier are cured.[360] Both could be assumed to be clinically appropriate and previous analyses have shown that predicted mortality reductions depend on the model used.[289 361] Future research could explore the use of another mechanism of screening benefit with the UK data. Future research could also explore a simpler version of the modelled natural history, for example merging T1-2 and T3-4 states into a pre-metastatic state, to determine the impact on results.

6.5. Conclusions

There is evidence that PSA testing is being undertaken opportunistically.[286] As this is likely to continue, it is important for policymakers to consider the optimal approach for screening. This analysis has used data from the ten-year follow-up of the CAP trial, amongst other sources, to show that a once-off screen at age 50 using PSA combined with a pre-biopsy MRI has the potential to be clinically effective and cost-effective in a UK setting. Uncertainty around this decision based on the parameter inputs has been comprehensively explored. Further exploration around appropriate utility values to be assigned to prostate cancer health states and the impact on results of changes in model structure is warranted.

CHAPTER 7. DISCUSSION

7.1. Introduction

The aim of this dissertation was to investigate methods for modelling the cost-effectiveness of screening interventions in an uncertain and rapidly changing landscape. As discussed throughout the chapters, cost-effectiveness models for screening interventions include many components that have the potential to change, including the population to screen, frequency of screening, screening test, diagnostic test or tests used, and their order and combination.[1] There may also be changing understanding with regard to the natural history of the condition in question and how this is impacted by the addition of new screening strategies or treatments. The methods explored to deal with this included:

(1) conducting a systematic review to explore uncertainty in previous model structures, parameter inputs, and the cost-effectiveness of novel tests and diagnostic strategies;

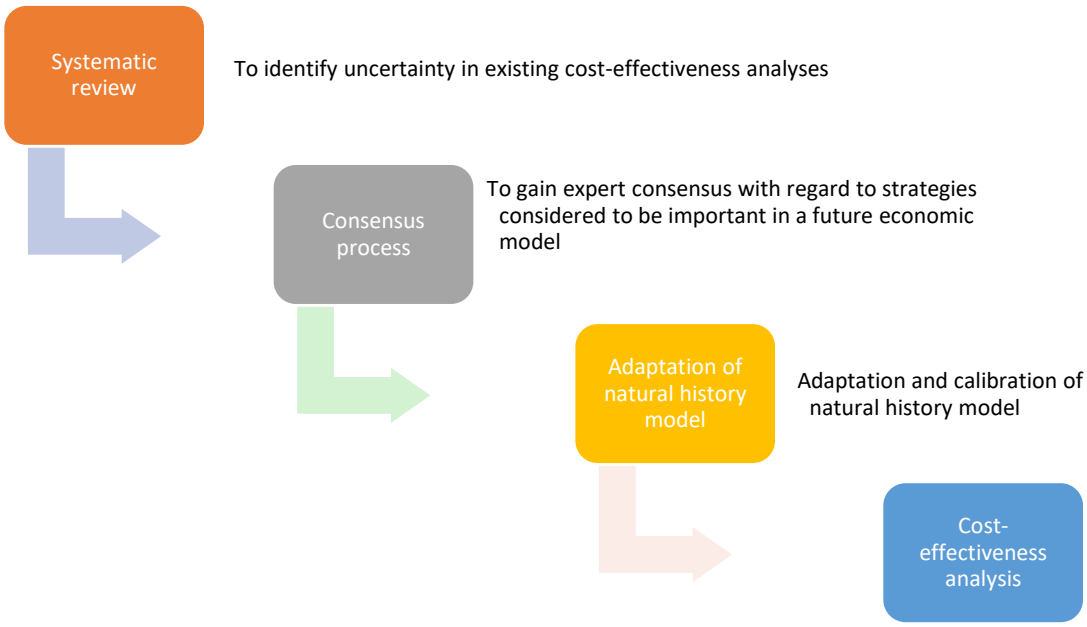
(2) determining which screening strategies to compare by gathering consensus views from experts on the screening strategies they considered to be relevant;

(3) adapting and calibrating a previously developed natural history model to a new setting; and finally,

(4) using the calibrated model to compare the cost-effectiveness of the strategies identified as relevant by the experts.

This process is shown in Figure 29.

Figure 29. Steps to deal with uncertainty in cost-effectiveness modelling of screening interventions



This chapter will summarise the overall findings of the dissertation before reflecting on the methods used and considering how this work compares with previous studies. Suggestions for areas of future research will also be provided.

7.2. Summary of findings

7.2.1. Step 1. Identifying uncertainty in existing cost-effectiveness models

The first method used to address uncertainty in the cost-effectiveness analysis of screening interventions was to conduct a systematic review of existing cost-effectiveness models to explore uncertainty in model structure and parameter inputs, including how disease progression is modelled. This is a common first step in the development of health economic models.[362] The aim of the systematic review was to identify cost-effectiveness models evaluating new diagnostic tests for prostate cancer to aid in the development of a new model. Twenty-two published studies were identified between 2011 and 2021. The review highlighted the need to ensure disease progression in diagnosed and undiagnosed cases is accurately represented and uncertainty is fully accounted for. It helped to determine the

strengths and limitations of previous models and current understanding around natural history. It also identified an existing model, the Prostata model[264], that was then adapted to answer the decision question specified in Chapter 6.

7.2.2. Step 2. Determining which screening strategies to compare

The second method involved gathering consensus views from experts on relevant screening strategies, given recent advancements. The modified-Delphi method was used. Views were elicited from 20 experts working in the field, over two rounds of online questionnaires, and agreement was obtained on relevant patient characteristics and screening technologies. The screening strategies considered relevant were:

1. No screening
2. Inviting all men within a certain age range to be screened
3. Inviting only higher risk men for screening
4. Inviting all men within a certain age bracket for screening but screening higher risk men at an earlier age
5. Using different screening intervals for higher and lower risk men
6. Using different screening intervals based on PSA level at a previous test

The panel did not reach consensus on exact age ranges to screen or a specific screening interval, with comments suggesting that these should be risk-based. However, limits were suggested by the participants which were then explored in the cost-effectiveness analysis described in step 4.

7.2.3. Step 3. Adapting and calibrating a model

The next step involved adapting the Prostata model, identified in the systematic review as a comprehensive and detailed natural history model, to allow the cost-effectiveness analysis of the screening strategies identified in the consensus process. The code for this model was available open source, allowing a more straightforward adaptation. The model was calibrated to a UK setting using national prostate cancer incidence data from the ONS by age[285], and data from the CAP trial on prostate cancer incidence by age and Gleason grade.[23] The calibration to the model's rate of prostate cancer onset parameter was carried out over two steps and used a Poisson likelihood with the BOBYQA algorithm.[290] Other calibration

methods were trialled, but none resulted in improved fit of the model's predictions to the observed data. The calibrated model showed good prediction of prostate cancer incidence in the UK on visual inspection of observed and predicted prostate cancer cumulative incidence rates, and a validation exercise gave mortality rate ratios that were broadly consistent with the CAP[23] and ERSPC[65] trials. This indicated that the natural history model was valid and could give reasonable predictions of the impact of introducing screening strategies in the UK. The use of UK-specific data on other parameters such as costs, treatment allocation, and mortality additionally prepared the model for use in a UK setting.

7.2.4. Step 4: Cost-effectiveness analysis

The cost-effectiveness analysis was then carried out using the adapted Prostata model and simulating the life histories of 10 million men under a range of screening strategies, based on those identified in step 2. The analysis estimated the lifetime cost-effectiveness, in terms of cost per QALYs gained, of prostate cancer screening from the perspective of the UK NHS. Other outcomes such as overdiagnosis and numbers of biopsies were also compared. The base case results showed that, of the strategies compared, a once-off PSA screen at 50 years old was the only strategy which would currently be considered cost-effective in the UK with an ICER of £12,860 per QALY gained, compared to no screening. This strategy resulted in only 2 overdiagnosed cases per 10,000 men compared to 113 overdiagnosed cases per 10,000 men in a strategy where screening is repeated every 2 years. Sensitivity analysis using disease-specific utility values suggested that adaptive screening (where men with PSA levels of < 1.5 ng/ml are screened every 6 years and values of > 1.5 ng/ml are screened every 4 years) may also be cost-effective. These are noteworthy findings given the current recommendation against prostate cancer screening in the UK.[355]

7.3. Reflection on methods used

The steps described in this dissertation provide a guide to future analysts hoping to undertake a cost-effectiveness analysis of screening interventions. The methods chosen were considered to be optimal for the application to prostate cancer screening and given

the time available; however, at each step alternative methods may have been used which are also worth considering.

7.3.1. Step 1. Identifying uncertainty in existing cost-effectiveness models

With regard to the systematic review, a limitation of using this method to inform the development of a new model is the time needed to complete the task, particularly if the literature is wide. Failing to assess the available literature may lead to the development of an inadequate model which does not appropriately take account of the natural history of the disease or use the best available evidence to inform model parameters. However, an alternative approach might involve a targeted review narrowed towards identifying key cost-effectiveness models. In some cases this may be more practical and an effectively targeted review may provide the majority of relevant information needed.[363]

7.3.2. Step 2. Determining which screening strategies to compare

With regard to the consensus process, an alternative method may have been to include all potential comparators in the cost-effectiveness analysis. This would be recommended in the case where there is an established diagnostic strategy e.g. a PSA higher than 3 ng/ml leads to a TRUS-guided biopsy, and the only question relates to screening interval or age ranges, for example. In the current scenario in prostate cancer screening, however, questions abound as to the diagnostic tests to use and the population eligible for screening, with innovations regularly changing the landscape. Although including a wide range of different intensities for each screening strategy would avoid bias and ensure no potentially cost-effective strategy was overlooked[239], it would have been computationally intensive in this instance, particularly considering the large amount of combinations of intervals, age ranges, tests and risk thresholds feasible. Improvements in software and modelling methods to decrease the computational time needed to simulate a screening population and produce

results may make this possible for a future analyst. In this case, however, this study has shown the benefit of expert input as to the strategies that may be deemed relevant.

The choice of comparators and comparator selection also depends on the aim of the analysis. The aim of the cost-effectiveness analysis conducted as part of this dissertation was to compare a selection of screening strategies that international researchers, clinicians and decision makers considered to be currently relevant, given the rapidly changing landscape, and identify the capacity for screening to be cost-effective in a UK setting. If the aim had been narrower, for example to compare the cost-effectiveness of all novel biomarkers as screening tests for prostate cancer or different age ranges to start and stop screening, a consensus process to identify screening strategies may not have been necessary.

7.3.3. Step 3. Adapting and calibrating a model

An alternative option would have been to develop a *de novo* model to compare the screening strategies. This would perhaps have been unnecessary in this instance given the substantial body of work on prostate cancer natural history modelling in the literature. Recycling and adapting previous models reduces duplication of effort and eases comparisons with previous modelling exercises. In a situation where the medical landscape is rapidly changing, as in prostate cancer screening, an analysts time may be better spent adapting previous models to enable comparison of newly relevant screening strategies, rather than 'reinventing the wheel' in terms of modelling natural history.[364-366]

The systematic review demonstrated the advantage in prostate cancer screening of having multiple previously developed models available to form the starting point for a new analysis. A particular advantage in this case is that the R code for a comprehensive and well-validated model (the Prostata model), shown to be capable of replicating observed outcomes from large trials, was available open-source on Github.[264] This avoided the need to rebuild the model based on published materials alone. Similar open source models have been identified for colorectal cancer screening[367] and breast cancer screening[368], and the literature demonstrates a general move towards the availability of open source models.[369-371]

A limitation is that the availability of an appropriate open-source model will likely not be the case in all indications and it has been shown that model replication based on information provided in publications alone can be a difficult task.[372] Model adaptation or replication can also be difficult without input or advice from the original model developer. An alternative approach would be to draw on previous models to inform model structure and parameters, while developing a *de novo* model. A benefit of developing a *de novo* model is that it can be more easily tailored to reflect current understanding around natural history. If it is found that previous models are outdated in their perception of how a cancer progresses, for example, the work involved in adapting a previous model to reflect current understanding may be considerable and may negate any benefits over developing a new model.

A degree of calibration is often necessary in screening models as it allows the estimation of parameter values which are not directly observable, such as the rates of clinical presentation.[113] In some screening indications, reliable data may not be available with which to estimate unobservable natural history parameters. An alternative approach may be to use expert elicitation to estimate the unobservable parameters. This approach was taken for the MISCAN colorectal model.[373] [374] There are several limitations to this approach, however, including biases in the sample group due to personal beliefs or experiences, lack of expertise to estimate complex parameters, communication challenges and the time taken to complete such an exercise.[375]

7.3.4. Step 4: Cost-effectiveness analysis

The cost-effectiveness analysis used the Prostata natural history model, without any changes to the underlying model assumptions, to compare the cost-effectiveness of the relevant screening strategies. Although parameter uncertainty was comprehensively assessed, a limitation of choosing one model to adapt, from the multiple models available, is that this does not take account of structural uncertainty and the impact of using different model structures and assumptions on cost-effectiveness results.[125] [376] [377] An alternative approach would take account of this uncertainty by conducting the analysis using a set of

different plausible models and using model averaging to produce results that have considered such structural uncertainty.[125 378]

7.4. Comparison with previous guides for carrying out cost-effectiveness analyses of screening interventions

Only one previous guide for the modelling of screening interventions was identified. The report by Karnon et al provides guidelines and good practice for model-based cost-utility analyses of screening programmes.[132] They do not make recommendations with regard to choosing comparators, other than excluding screening strategies that are not considered feasible. In accordance with the approach taken in this dissertation, they also recommend natural history modelling to take account of cancer progression and that a review of existing screening models should be carried out before deciding on model structure. Karnon et al agree that discrete event simulation models can overcome the limitations of cohort Markov models but warn that they may have significantly longer running times. Calibration of incidence rates to observed prevalence rates of different stages of disease is also recommended as well as the use of models which describe disease progression from the point at which the disease becomes detectable to death. In comparison to the guidelines provided by Karnon et al, this dissertation offers more of a practical guide to carrying out cost-effectiveness analyses of screening interventions, proposing step by step methods and demonstrating these with an applied example.

NICE provides guidance on the cost-effectiveness analysis of diagnostic tests and technologies, including screening tests, in their Diagnostic Assessment Programme manual.[1] This states that the comparators in a cost-effectiveness analysis should be the technologies or tests that are most commonly used or are recommended in current NICE guidance. No specific model type or method to estimate parameters is recommended but it is stated that existing models can be used as an alternative to de novo modelling if they are 'adequate

and appropriate'. They recommend that a systematic search for existing models of cost effectiveness may not be necessary as the objective is to identify any appropriate and high-quality models, rather than all models.

7.5. Comparison with methods used in previous cancer screening studies

7.5.1. How previous prostate cancer screening models decided on model structure

Of the 22 studies included in the systematic review in Chapter 3, 13 developed a *de novo* model and did not discuss how the model was informed. Two studies carried out a systematic review but developed a *de novo* model as no cost-effectiveness models relevant to their setting could be identified.[157 379] Three developed a *de novo* model and stated that guidelines, expert opinion and data from trials informed their model development.[140 155 163] Only four studies took the same approach as in this study, adapting a previously developed model, but did not state how the model was chosen.[143 148 150 151] As mentioned, in the area of prostate cancer screening, continuously developing *de novo* models each time a new decision question needs to be answered may be unnecessary and time may be better spent ensuring that models that are developed are available to others, easily understandable and adaptable.[364-366]

7.5.2. How previous prostate cancer screening models chose comparators

In previous analyses it is often unclear how comparators have been chosen. Of the 22 studies included in the systematic review, for example, 13 did not provide any rationale for the comparators included. Of those that did, 3 provided a justification for screening frequency only[28 150 151], 1 provided a justification for screening age-ranges only[149], and 1 provided a justification only for screening frequency and age ranges considered[143], with all basing these on government recommendations or trial protocols. One compared all clinically

feasible combinations of tests considered but did not consider different age ranges or screening intervals[152], and two based the comparators chosen on the availability of data.[163 379] Only one study stated that the comparators chosen were based on those an expert committee deemed clinically meaningful, although the process to determine meaningfulness was not described.[157]

7.5.3. Calibration in previous prostate cancer screening models

Only four of the cost-effectiveness models identified in the systematic review used calibration to determine model parameters.[143 148 152 157] Two of these were based on analyses carried out by the CISNET group, including investigators from the Fred Hutchinson Cancer Research Center in the US and the Erasmus University Medical Center in the Netherlands, who have used the same core natural history models to assess the cost-effectiveness of various prostate cancer screening strategies, in multiple settings.[28 293] The Prostata model used in this analysis, has also been re-calibrated several times for different decision questions in different settings including whether the use of MRI and/or the Stockholm3 test is cost-effective in prostate cancer screening in Sweden.[143 284 295] As mentioned in section 2.5.1, without calibration to relevant data sources, it is difficult to estimate the unobservable natural history parameters which are critical in models of disease progression.

7.6. Strengths and limitations of dissertation overall

As a guide to carrying out cost-effectiveness analysis of screening interventions, the strengths of this dissertation are the detailed description of methods with a practical application. This dissertation has provided an assessment of the potential impact of relevant screening strategies in the UK today, with a wide range of sensitivity and scenario analyses used to explore the robustness of the results. Twenty experts in the field of prostate cancer screening were involved in the decision on which screening strategies were relevant to compare in the cost-effectiveness model. It is also the first UK study to consider prostate

cancer progression when modelling the cost-effectiveness of the use of polygenic risk scores within a screening programme.

Another strength of this dissertation is that it has built on a large body of previous work in the area of prostate cancer natural history modelling by adapting and calibrating a previously developed natural history model. It has also made best use of rich individual patient data from a large UK data source (the CAP trial), to inform the updated UK-based modelling. In doing so, it has provided a well-calibrated UK-specific natural history model which can be used in future analyses to update cost-effectiveness recommendations as new data on screening interventions become available.

Limitations are that not all potential methods for modelling screening interventions in an uncertain landscape have been explored, particularly methods to take account of different potential model structures and characterisations of disease progression. This was not feasible due to time constraints but is a key area for future development of the model. The use of expert opinion to decide on relevant screening strategies comes with the limitation that the sample of experts chosen may have had some bias towards particular screening methods. However, this is a step beyond common practice in cost-effectiveness modelling where the comparators are normally chosen by the study team alone (section 7.5.2).

A general limitation of using calibration to estimate model parameters and achieve a model whose outputs match with observed data is that one must be confident that the calibration targets used are representative of the population considered in the decision question. Although the CAP trial, due to its long-term follow-up (over 10 years), can provide much information on how cancers progress over time, as it was initiated in the year 2000, the data is limited by the fact that diagnosis and management have now changed. Future work could adjust for the fact that cancers diagnosed in the CAP trial may be staged differently if diagnosed today.

A further limitation is that several aspects of the cost-effectiveness modelling may now be out of date. Treatment allocation, for example, was based on 2016 data from the National Cancer Registration and Analysis Service in the UK.[298] It is likely that current treatment patterns have changed or are changing, with a move towards increased active surveillance of low-risk cancers.[157] This is again an area for future exploration.

7.7. Recommendations for further research

Throughout the chapters, recommendations for further research have been made including updating the model as new data and new screening and diagnostic practices emerge and considering the impact of different model structures. As there is interest in risk-stratified screening, recommendations have also been made to explore this space in more detail including considering the use of different screening strategies in lower and higher risk groups and exploring the use of different risk thresholds. The analysis reported in this dissertation considered only risk-stratification by polygenic risk score, but the consensus process highlighted that there is also interest in stratification by other factors including ethnicity, family history and life expectancy. In addition, the analysis did not take account of barriers to implementing polygenic risk-stratified screening, assuming that genetic samples for all men would be available at a low cost. Methods are available to quantify the impact of capacity constraints that may exist here such as limited testing capacity and a reluctance in men to participate.[380]

The consensus process highlighted the value of expert opinion when developing a cost-effectiveness model. Future analysts could consider using a similar process to determine, not only relevant comparators, but other model aspects such as relevant health states and key natural history parameters.[381]

As it is hoped that this dissertation will provide a guide to analysts undertaking a new cost-effectiveness analysis of screening interventions, the overall recommendation is to consider the steps taken and report on their utility and relevance in other indications. This in turn will help future analysts to refine recommendations with regard to such analyses.

7.8. Conclusions

The aim of this PhD was to provide a guide to identifying and dealing with the uncertainty that is inherent in cost-effectiveness modelling of screening strategies, using prostate cancer screening as a case study. To meet this aim, a systematic literature review was carried out to assess the evidence base on recent cost-effectiveness models which have considered new innovations in prostate cancer screening. This review helped to inform the baseline for the next step in the process, which was the use of the modified-Delphi method to gain consensus on relevant screening and diagnostic strategies to compare in a future cost-effectiveness analysis. This involved 20 experts including clinicians, modellers, experts in prostate cancer and other relevant stakeholders. The systematic review also helped to identify an openly available natural history model that, in the next step, was adapted and calibrated for use in a UK setting. Many aspects of the model were adapted including assumptions around background PSA testing, treatment allocation and costs and utilities. The model was then calibrated to data from over 400,000 men who participated in the CAP trial, as well as registry data from the UK ONS. The final step was to use the agreed strategies from the modified-Delphi process and the calibrated model to carry out a cost-effectiveness analysis of prostate cancer screening strategies in the UK based on the outcome of cost per QALYs gained. This final chapter has reflected on the methods used and made recommendations for future analysts.

The key contributions of this dissertation are that it provides a practical guide to the cost-effectiveness modelling of screening interventions under conditions of uncertainty. By applying the methods explored in the area of prostate cancer screening, it has determined the cost-effectiveness of various prostate cancer screening strategies in a UK setting, including risk-stratified and adaptive approaches. The conclusion that a once-off screen at age 50 has the potential to be effective and cost-effective is useful to decision makers, particularly amidst ongoing uncertainty as to whether screening should be provided in the UK.[355] These results were directly considered by the UK National Screening Committee at a workshop organised to discuss the future of prostate cancer screening in June 2021. Finally,

this work has provided a UK-adapted and calibrated model that can be used for future research.

The methods explored are a few of many possible methods available to deal with the many types of uncertainty associated with modelling of cancer screening interventions. The recommendations come with the limitations of the availability of previous models to adapt and data by which to update the selected model. Nevertheless, the steps described provide a good starting point for any analyst hoping to undertake a future cost-effectiveness analysis of novel screening strategies. It is hoped that the use of this guide could lead to an improvement in the quality of cost-effectiveness models published in the area of screening.

Bibliography

1. NICE. NICE Process and Methods Guides. Diagnostics Assessment Programme Manual. London: National Institute for Health and Care Excellence (NICE) 2011.
2. Wilson JMG, Jungner G, Organization WH. Principles and practice of screening for disease. 1968
3. Lee JM, Lehman CD, Miglioretti DL. Screening trials and design. Handbook for Clinical Trials of Imaging and Image-Guided Interventions2016:76-90.
4. Criteria for a population screening programme: UK National Screening Committee; 2022 [Available from: https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme.
5. Davies-Oliveira JC, Round T, Crosbie EJ. Cervical screening: the evolving landscape. *British Journal of General Practice* 2022;72(721):364. doi: 10.3399/bjgp22X720197
6. Allweis TM, Hermann N, Berenstein-Molho R, et al. Personalized Screening for Breast Cancer: Rationale, Present Practices, and Future Directions. *Annals of Surgical Oncology* 2021;28(8):4306-17. doi: 10.1245/s10434-020-09426-1
7. Joshi S. Common Clinical Trial Amendments, why they are submitted and how they can be avoided: a mixed methods study on NHS UK Sponsored Research (Amendments Assemble). *Trials* 2023;24(1):10. doi: 10.1186/s13063-022-06989-0
8. Kernick DP. Introduction to health economics for the medical practitioner. *Postgraduate Medical Journal* 2003;79(929):147. doi: 10.1136/pmj.79.929.147
9. McPake B, Normand C, Smith S, et al. Health economics: an international perspective: Routledge 2020.
10. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation: Oup Oxford 2006.
11. Bespalov A, Barchuk A, Auvinen A, et al. Cancer screening simulation models: a state of the art review. *BMC Medical Informatics and Decision Making* 2021;21(1):359. doi: 10.1186/s12911-021-01713-5
12. Philips Z, Bojke L, Sculpher M, et al. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment. *PharmacoEconomics* 2006;24(4):355-71. doi: 10.2165/00019053-200624040-00006
13. Getaneh AM, Heijnsdijk EAM, Roobol MJ, et al. Assessment of harms, benefits, and cost-effectiveness of prostate cancer screening: A micro-simulation study of 230 scenarios. *Cancer Med* 2020;9(20):7742-50. doi: 10.1002/cam4.3395 [published Online First: 2020/08/20]
14. Saumoy M, Schneider Y, Shen N, et al. Cost Effectiveness of Gastric Cancer Screening According to Race and Ethnicity. *Gastroenterology* 2018;155(3):648-60. doi: http://dx.doi.org/10.1053/j.gastro.2018.05.026
15. Stout NK, Knudsen AB, Kong CY, et al. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics* 2009;27(7):533-45. doi: 10.2165/11314830-000000000-00000 [published Online First: 2009/08/12]
16. CRUK. Prostate Cancer Statistics 2022 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer accessed 4/10/22.
17. Nafissi NN, Kosiorek HE, Butterfield RJ, et al. Evolving Natural History of Metastatic Prostate Cancer. *Cureus* 2020;12(11):e11484. doi: 10.7759/cureus.11484

18. Crawford ED. Understanding the Epidemiology, Natural History, and Key Pathways Involved in Prostate Cancer. *Urology* 2009;73(5):S4-S10. doi: 10.1016/j.urology.2009.03.001
19. Heijnsdijk EAM, de Koning HJ. Improved Harm/Benefit Ratio and Cost-effectiveness of Prostate Cancer Screening Using New Technologies. *European Urology* 2022;82(1):20-21. doi: 10.1016/j.eururo.2022.01.035
20. Hamdy FC, Donovan JL, Lane JA, et al. Fifteen-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *New England Journal of Medicine* 2023 doi: 10.1056/NEJMoa2214122
21. Schröder FH, Hugosson J, Roobol MJ, 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 (London, England)* 2014;384(9959):2027-35. doi: 10.1016/S0140-6736(14)60525-0 [published Online First: 2014/08/06]
22. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute* 2012;104(2):125-32. doi: 10.1093/jnci/djr500 [published Online First: 2012/01/06]
23. Martin RM, Donovan JL, Turner EL, et al. Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial. *JAMA* 2018;319(9):883-95. doi: 10.1001/jama.2018.0154
24. Sanghera S, Coast J, Martin RM, et al. Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models. 2018;18(1):84. doi: 10.1186/s12885-017-3974-1
25. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet* 2017;389(10071):815-22. doi: https://doi.org/10.1016/S0140-6736(16)32401-1
26. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. *JNCI: Journal of the National Cancer Institute* 2009;101(6):374-83. doi: 10.1093/jnci/djp001
27. Pashayan N, Duffy SW, Neal DE, et al. Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genetics in Medicine* 2015;17(10):789-95.
28. Callender T, Emberton M, Morris S, et al. Benefit, Harm, and Cost-effectiveness Associated With Magnetic Resonance Imaging Before Biopsy in Age-based and Risk-stratified Screening for Prostate Cancer. *JAMA Network Open* 2021;4(3):e2037657-e57. doi: 10.1001/jamanetworkopen.2020.37657
29. Andrew J Martin SJL, Hannah E Verry, Martin R Stockler and Jon D Emery. Risk assessment to guide prostate cancer screening decisions: a cost-effectiveness analysis. *The Medical Journal of Australia* 2013;198(10):546-50. doi: 10.5694/mja12.11597
30. Heijnsdijk EA, De Carvalho T, Auvinen A, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. 2015;107(1)
31. Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm: World Health Organization; 2020 [Available from: https://iris.who.int/bitstream/handle/10665/330829/9789289054782-eng.pdf?sequence=1.

32. Swift A, Heale R, Twycross A. What are sensitivity and specificity? *Evidence Based Nursing* 2020;23(1):2. doi: 10.1136/ebnurs-2019-103225
33. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *Bmj* 1994;309(6947):102. doi: 10.1136/bmj.309.6947.102 [published Online First: 1994/07/09]
34. Feldman W. How serious are the adverse effects of screening? *J Gen Intern Med* 1990;5(5 Suppl):S50-3. doi: 10.1007/bf02600842 [published Online First: 1990/09/01]
35. Brodersen J, Schwartz LM, Heneghan C, et al. Overdiagnosis: what it is and what it isn't. *BMJ Evidence-Based Medicine* 2018;23(1):1. doi: 10.1136/ebmed-2017-110886
36. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. 2014;12(1):26.
37. What is Prostate Cancer? : Centers for Disease Control and Prevention; 2023 [Available from: https://www.cdc.gov/cancer/prostate/basic_info/what-is-prostate-cancer.htm.
38. How does the prostate work? : Institute for Quality and Efficiency in Health Care; 2006 [Available from: https://www.ncbi.nlm.nih.gov/books/NBK279291/.
39. Grossman DC, Curry SJ, Owens DK, et al. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* 2018;319(18):1901-13. doi: 10.1001/jama.2018.3710 [published Online First: 2018/05/26]
40. Hamilton W, Sharp D. Symptomatic diagnosis of prostate cancer in primary care: a structured review. *British Journal of General Practice* 2004;54(505):617.
41. Schröder F, Hermanek P, Denis L, et al. The TNM classification of prostate cancer. *The Prostate* 1992;21(S4):129-38.
42. CRUK. Grade groups of prostate cancer 2023 [Available from: https://www.cancerresearchuk.org/about-cancer/prostate-cancer/stages/grades accessed 5/4/23.
43. Moch H, Amin MB, Berney DM, et al. The 2022 World Health Organization Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *European urology* 2022
44. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *The American journal of surgical pathology* 2016;40(2):244-52.
45. CRUK. Prostate cancer survival 2023 [Available from: https://www.cancerresearchuk.org/about-cancer/prostate-cancer/survival#:~:text=Survival%20for%20all%20stages%20of%20prostate%20cancer&text=more%20than%2095%20out%20of,for%2010%20years%20or%20more.
46. ASCO Answers: Prostate cancer: AMERICAN SOCIETY OF CLINICAL ONCOLOGY; 2021 [Available from: https://www.cancer.net/sites/cancer.net/files/asco_answers_prostate.pdf.
47. Prostate cancer statistics: Prostate Cancer UK; 2023 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-Two.
48. Lloyd T, Hounsome L, Mehay A, et al. Lifetime risk of being diagnosed with, or dying from, prostate cancer by major ethnic group in England 2008–2010. *BMC Medicine* 2015;13(1):171. doi: 10.1186/s12916-015-0405-5

49. Kiciński M, Vangronsveld J, Nawrot TS. An Epidemiological Reappraisal of the Familial Aggregation of Prostate Cancer: A Meta-Analysis. *PLOS ONE* 2011;6(10):e27130. doi: 10.1371/journal.pone.0027130
50. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics* 2018;50(7):928-36. doi: 10.1038/s41588-018-0142-8
51. Conti DV, Darst BF, Moss LC, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature genetics* 2021;53(1):65-75.
52. Matejcic M, Saunders EJ, Dadaev T, et al. Germline variation at 8q24 and prostate cancer risk in men of European ancestry. *Nature communications* 2018;9(1):1-11.
53. Dadaev T, Saunders EJ, Newcombe PJ, et al. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nature Communications* 2018;9(1):2256. doi: 10.1038/s41467-018-04109-8
54. Roland M, Neal D, Buckley R. What should doctors say to men asking for a PSA test? *BMJ* 2018;362:k3702. doi: 10.1136/bmj.k3702
55. Merriel SWD, Pocock L, Gilbert E, et al. Systematic review and meta-analysis of the diagnostic accuracy of prostate-specific antigen (PSA) for the detection of prostate cancer in symptomatic patients. *BMC Medicine* 2022;20(1):54. doi: 10.1186/s12916-021-02230-y
56. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized european study. *New England Journal of Medicine* 2009;360(13):1320-28. doi: 10.1056/NEJMoa0810084
57. Rosario DJ, Lane JA, Metcalfe C, et al. Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. 2012;344 doi: 10.1136/bmj.d7894 %J BMJ
58. Roehl KA, Antenor JAV, Catalona WJ. Serial Biopsy Results in Prostate Cancer Screening Study. *The Journal of Urology* 2002;167(6):2435-39. doi: https://doi.org/10.1016/S0022-5347(05)64999-3
59. Van Poppel H, Roobol MJ, Chapple CR, et al. Prostate-specific Antigen Testing as Part of a Risk-Adapted Early Detection Strategy for Prostate Cancer: European Association of Urology Position and Recommendations for 2021. *European Urology* 2021;80(6):703-11. doi: 10.1016/j.eururo.2021.07.024
60. Jackson SD, de la Rue MR, Greenslade TP, et al. Screening asymptomatic men for prostate cancer: A comparison of international guidelines on prostate-specific antigen testing. *J Med Screen* 2022;29(4):268-71. doi: 10.1177/09691413221119238 [published Online First: 20220904]
61. European Health Union: a new EU approach on cancer detection – screening more and screening better. : European Commission; 2022 [Available from: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_5562.
62. NHS. PSA testing 2023 [Available from: https://www.nhs.uk/conditions/prostate-cancer/psa-testing/ accessed 4/4/23.
63. Smith RA, Andrews KS, Brooks D, et al. Cancer screening in the United States, 2018: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2018;68(4):297-316. doi: 10.3322/caac.21446 [published Online First: 2018/05/31]

64. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. *The Journal of urology* 2013;190(2):419-26.
65. Hugosson J, Roobol MJ, Månsson M, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol* 2019;76(1):43-51. doi: 10.1016/j.eururo.2019.02.009 [published Online First: 2019/03/03]
66. Saarimäki L, Hugosson J, Tammela TL, et al. Impact of Prostatic-specific Antigen Threshold and Screening Interval in Prostate Cancer Screening Outcomes: Comparing the Swedish and Finnish European Randomised Study of Screening for Prostate Cancer Centres. *Eur Urol Focus* 2019;5(2):186-91. doi: 10.1016/j.euf.2017.07.007 [published Online First: 20170810]
67. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and Prostate-Cancer Mortality in a Randomized European Study. *New England Journal of Medicine* 2009;360(13):1320-28. doi: 10.1056/NEJMoa0810084
68. Labrie F, Candas B, Cusan L, et al. Screening decreases prostate cancer mortality: 11-year follow-up of the 1988 Quebec prospective randomized controlled trial. 2004;59(3):311-18. doi: doi:10.1002/pros.20017
69. Martin RM, Donovan JL, Turner EL, et al. Effect of a low-intensity psa-based screening intervention on prostate cancer mortality: The cap randomized clinical trial. *JAMA* 2018;319(9):883-95. doi: 10.1001/jama.2018.0154
70. Pinsky PF, Prorok PC, Yu K, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. *Cancer* 2017;123(4):592-99. doi: 10.1002/cncr.30474 [published Online First: 2016/12/03]
71. Lundgren P-O, Kjellman A, Norming U, et al. Long-Term Outcome of a Single Intervention Population Based Prostate Cancer Screening Study. *Journal of Urology* 2018;200(1):82-88. doi: 10.1016/j.juro.2018.01.080
72. Carter HB, Piantadosi S, Isaacs JT. Clinical evidence for and implications of the multistep development of prostate cancer. *The Journal of urology* 1990;143(4):742-46.
73. Etzioni R, Cha R, Feuer EJ, et al. Asymptomatic Incidence and Duration of Prostate Cancer. *American Journal of Epidemiology* 1998;148(8):775-85. doi: 10.1093/oxfordjournals.aje.a009698
74. Srivastava S, Koay EJ, Borowsky AD, et al. Cancer overdiagnosis: a biological challenge and clinical dilemma. *Nat Rev Cancer* 2019;19(6):349-58. doi: 10.1038/s41568-019-0142-8
75. Sipeky C, Talala KM, Tammela TLJ, et al. Prostate cancer risk prediction using a polygenic risk score. *Sci Rep* 2020;10(1):17075-75. doi: 10.1038/s41598-020-74172-z
76. Chowdhury S, Dent T, Pashayan N, et al. Incorporating genomics into breast and prostate cancer screening: assessing the implications. *Genetics in Medicine* 2013;15(6):423-32.
77. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine* 2020;12(1):44. doi: 10.1186/s13073-020-00742-5
78. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *New England Journal of Medicine* 2018;378(19):1767-77. doi: 10.1056/NEJMoa1801993
79. Eklund M, Jäderling F, Discacciati A, et al. MRI-Targeted or Standard Biopsy in Prostate Cancer Screening. *New England Journal of Medicine* 2021 doi: 10.1056/NEJMoa2100852

80. Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the economic evaluation of health care programmes: Oxford university press 2015.
81. Williams A. Priority setting in public and private health care: a guide through the ideological jungle. *Journal of health economics* 1988;7(2):173-83.
82. Shiell A, Donaldson C, Mitton C, et al. Health economic evaluation. *Journal of Epidemiology and Community Health* 2002;56(2):85. doi: 10.1136/jech.56.2.85
83. Bernfort L, Gerdle B, Husberg M, et al. People in states worse than dead according to the EQ-5D UK value set: would they rather be dead? *Qual Life Res* 2018;27(7):1827-33. doi: 10.1007/s11136-018-1848-x [published Online First: 2018/04/05]
84. Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin* 2010;96(1):5-21. doi: 10.1093/bmb/ldq033
85. Brooks R, Group E. EuroQol: the current state of play. *Health policy* 1996;37(1):53-72.
86. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of health economics* 2002;21(2):271-92.
87. Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health and quality of life outcomes* 2003;1(1):1-13.
88. Brazier J, Dixon S. The use of condition specific outcome measures in economic appraisal. *Health Economics* 1995;4(4):255-64.
89. NICE. Guide to the methods of technology appraisal 2013 2013 [Available from: https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781.
90. Robinson R. Cost-benefit analysis. *Bmj* 1993;307(6909):924-6. doi: 10.1136/bmj.307.6909.924 [published Online First: 1993/10/09]
91. Seixas BV. Welfarism and extra-welfarism: a critical overview. *Cad Saude Publica* 2017;33(8):e00014317. doi: 10.1590/0102-311x00014317 [published Online First: 2017/08/24]
92. Brouwer WBF, Culyer AJ, van Exel NJA, et al. Welfarism vs. extra-welfarism. *Journal of Health Economics* 2008;27(2):325-38. doi: https://doi.org/10.1016/j.jhealeco.2007.07.003
93. Turner HC, Archer RA, Downey LE, et al. An Introduction to the Main Types of Economic Evaluations Used for Informing Priority Setting and Resource Allocation in Healthcare: Key Features, Uses, and Limitations. *Front Public Health* 2021;9:722927. doi: 10.3389/fpubh.2021.722927 [published Online First: 2021/09/14]
94. Webb J, Fife-Schaw C, Ogden J. A randomised control trial and cost-consequence analysis to examine the effects of a print-based intervention supported by internet tools on the physical activity of UK cancer survivors. *Public Health* 2019;171:106-15. doi: https://doi.org/10.1016/j.puhe.2019.04.006
95. Byford S, Raftery J. Perspectives in economic evaluation. *Bmj* 1998;316(7143):1529-30. doi: 10.1136/bmj.316.7143.1529 [published Online First: 1998/06/06]
96. Dominance [online]: York Health Economics Consortium; 2016 [Available from: https://yhec.co.uk/glossary/dominance/.
97. Thokala P, Ochalek J, Leech AA, et al. Cost-Effectiveness Thresholds: the Past, the Present and the Future. *PharmacoEconomics* 2018;36(5):509-22. doi: 10.1007/s40273-017-0606-1
98. NICE health technology evaluations: the manual-Process and methods: National Institute for Health and Care Excellence 2022 [Available from:

https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741.
99. Cameron D, Ubels J, Norström F. On what basis are medical cost-effectiveness thresholds set? Clashing opinions and an absence of data: a systematic review. *Global Health Action* 2018;11(1):1447828. doi: 10.1080/16549716.2018.1447828
100. Claxton K, Martin S, Soares M, et al. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess* 2015;19(14):1-503, v-vi. doi: 10.3310/hta19140 [published Online First: 2015/02/19]
101. Paulden M. Calculating and Interpreting ICERs and Net Benefit. *PharmacoEconomics* 2020;38(8):785-807. doi: 10.1007/s40273-020-00914-6
102. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health economics* 2001;10(8):779-87.
103. Barton GR, Briggs AH, Fenwick EA. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfection information (EVPI). *Value Health* 2008;11(5):886-97. doi: 10.1111/j.1524-4733.2008.00358.x [published Online First: 20080516]
104. Sakowsky RA. Disentangling the welfarism/extra-welfarism distinction: Towards a more fine-grained categorization. *Health Economics* 2021;30(9):2307-11. doi: https://doi.org/10.1002/hec.4382
105. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 2006;15(12):1295-310. doi: https://doi.org/10.1002/hec.1148
106. Karnon J, Brown J. Selecting a decision model for economic evaluation: a case study and review. *Health Care Management Science* 1998;1(2):133-40. doi: 10.1023/A:1019090401655
107. Krijkamp EM, Alarid-Escudero F, Enns EA, et al. Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making* 2018;38(3):400-22. doi: 10.1177/0272989x18754513
108. Karnon J, Haji Ali Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics* 2014;32(6):547-58. doi: 10.1007/s40273-014-0147-9 [published Online First: 2014/03/15]
109. Siebert U, Alagoz O, Bayoumi AM, et al. State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–3. *Medical Decision Making* 2012;32(5):690-700. doi: 10.1177/0272989X12455463
110. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4. *Value Health* 2012;15(6):821-7. doi: 10.1016/j.jval.2012.04.013 [published Online First: 2012/09/25]
111. Vanni T, Karnon J, Madan J, et al. Calibrating Models in Economic Evaluation. *PharmacoEconomics* 2011;29(1):35-49. doi: 10.2165/11584600-000000000-00000
112. Banks J. Handbook of simulation: principles, methodology, advances, applications, and practice: John Wiley & Sons 1998.

113. Weinstein MC. Recent developments in decision-analytic modelling for economic evaluation. *Pharmacoeconomics* 2006;24(11):1043-53. doi: 10.2165/00019053-200624110-00002 [published Online First: 2006/10/28]
114. Kim JJ, Kuntz KM, Stout NK, et al. Multiparameter Calibration of a Natural History Model of Cervical Cancer. *American Journal of Epidemiology* 2007;166(2):137-50. doi: 10.1093/aje/kwm086
115. Van de Velde N, Brisson M, Boily MC. Modeling human papillomavirus vaccine effectiveness: quantifying the impact of parameter uncertainty. *Am J Epidemiol* 2007;165(7):762-75. doi: 10.1093/aje/kwk059 [published Online First: 2007/02/06]
116. Eddy DM, Hollingworth W, Caro JJ, et al. Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–7. *Medical Decision Making* 2012;32(5):733-43. doi: 10.1177/0272989X12454579
117. Karnon J, Vanni T. Calibrating Models in Economic Evaluation. *PharmacoEconomics* 2011;29(1):51-62. doi: 10.2165/11584610-000000000-00000
118. Nelder JA, Mead R. A simplex method for function minimization. *The computer journal* 1965;7(4):308-13.
119. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Med Decis Making* 2011;31(4):625-41. doi: 10.1177/0272989x10384738 [published Online First: 2010/12/04]
120. Amaran S, Sahinidis NV, Sharda B, et al. Simulation optimization: a review of algorithms and applications. *Annals of Operations Research* 2016;240(1):351-80. doi: 10.1007/s10479-015-2019-x
121. Enns EA, Cipriano LE, Simons CT, et al. Identifying Best-Fitting Inputs in Health-Economic Model Calibration: A Pareto Frontier Approach. *Medical Decision Making* 2014;35(2):170-82. doi: 10.1177/0272989X14528382
122. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health* 2012;15(6):835-42.
123. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2001;63(3):425-64. doi: https://doi.org/10.1111/1467-9868.00294
124. Chrysanthopoulou SA, Rutter CM, Gatsonis CA. Bayesian versus Empirical Calibration of Microsimulation Models: A Comparative Analysis. *Medical Decision Making* 2021;41(6):714-26. doi: 10.1177/0272989X211009161
125. Bojke L, Claxton K, Sculpher M, et al. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value Health* 2009;12(5):739-49. doi: 10.1111/j.1524-4733.2008.00502.x [published Online First: 2009/06/11]
126. Briggs AH. Handling Uncertainty in Cost-Effectiveness Models. *PharmacoEconomics* 2000;17(5):479-500. doi: 10.2165/00019053-200017050-00006
127. Tuffaha HW, Gordon LG, Scuffham PA. Value of information analysis in healthcare: a review of principles and applications. *Journal of Medical Economics* 2014;17(6):377-83. doi: 10.3111/13696998.2014.907170
128. Wilson ECF. A Practical Guide to Value of Information Analysis. *PharmacoEconomics* 2015;33(2):105-21. doi: 10.1007/s40273-014-0219-x
129. Newland A. NICE diagnostics assessment programme manual: The Royal College of Surgeons of England; 2011 [Available from:

https://www.nice.org.uk/media/default/about/what-we-do/nice-guidance/nice-diagnostics-guidance/diagnostics-assessment-programme-manual.pdf.
130. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions: John Wiley & Sons 2011.
131. Keeney E, Thom H, Turner E, et al. Systematic Review of Cost-Effectiveness Models in Prostate Cancer: Exploring New Developments in Testing and Diagnosis. *Value in Health* 2022;25(1):133-46. doi: https://doi.org/10.1016/j.jval.2021.07.002
132. Karnon J, Goyder E, Tappenden P, et al. A review and critique of modelling in prioritising and designing screening programmes. *Health Technol Assess* 2007;11(52)
133. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMJ : British Medical Journal* 2013;346:f1049. doi: 10.1136/bmj.f1049
134. Adarkwah CC, van Gils PF, Hiligsmann M, et al. Risk of bias in model-based economic evaluations: the ECOBIAS checklist. *Expert Rev Pharmacoecon Outcomes Res* 2016;16(4):513-23. doi: 10.1586/14737167.2015.1103185 [published Online First: 2015/11/21]
135. van Mastrigt GA, Hiligsmann M, Arts JJ, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). *Expert Rev Pharmacoecon Outcomes Res* 2016;16(6):689-704. doi: 10.1080/14737167.2016.1246960 [published Online First: 2016/11/03]
136. Thielen FW, Van Mastrigt G, Burgers LT, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). *Expert Rev Pharmacoecon Outcomes Res* 2016;16(6):705-21. doi: 10.1080/14737167.2016.1246962 [published Online First: 2016/11/03]
137. Wijnen B, Van Mastrigt G, Redekop WK, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). *Expert Rev Pharmacoecon Outcomes Res* 2016;16(6):723-32. doi: 10.1080/14737167.2016.1246961 [published Online First: 2016/10/21]
138. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71
139. Govers TM, Hessels D, Vlaeminck-Guillem V, et al. Cost-effectiveness of SelectMDx for prostate cancer in four European countries: a comparative modeling study. *Prostate Cancer & Prostatic Diseases* 2019;22(1):101-09. doi: https://dx.doi.org/10.1038/s41391-018-0076-3
140. Bouttell J, Teoh J, Chiu PK, et al. Economic evaluation of the introduction of the Prostate Health Index as a rule-out test to avoid unnecessary biopsies in men with prostate specific antigen levels of 4-10 in Hong Kong. *PLoS ONE [Electronic Resource]* 2019;14(4):e0215279. doi: https://dx.doi.org/10.1371/journal.pone.0215279
141. Schiffer E, Bick C, Grizelj B, et al. Urinary proteome analysis for prostate cancer diagnosis: Cost-effective application in routine clinical practice in Germany. 2012;19(2):118-25. doi: 10.1111/j.1442-2042.2011.02901.x
142. Kim L, Boxall N, George A, et al. Clinical utility and cost modelling of the phi test to triage referrals into image-based diagnostic services for suspected prostate cancer: the PRIM (Phi to RefIne Mri) study. *BMC Medicine* 2020;18(1):95. doi: 10.1186/s12916-020-01548-3

143. Karlsson AA, Hao S, Jauhiainen A, et al. The cost-effectiveness of prostate cancer screening using the Stockholm3 test. *PLoS One* 2021;16(2):e0246674. doi: 10.1371/journal.pone.0246674 [published Online First: 2021/02/26]
144. Teoh JY-C, Leung C-H, Wang MH, et al. The cost-effectiveness of prostate health index for prostate cancer detection in Chinese men. *Prostate Cancer and Prostatic Diseases* 2020;23(4):615-21. doi: 10.1038/s41391-020-0243-1
145. Govers TM, Caba L, Resnick MJ. Cost-Effectiveness of Urinary Biomarker Panel in Prostate Cancer Risk Assessment. *Journal of Urology* 2018;200(6):1221-26. doi: https://dx.doi.org/10.1016/j.juro.2018.07.034
146. Sathianathen NJ, Kuntz KM, Alarid-Escudero F, et al. Incorporating Biomarkers into the Primary Prostate Biopsy Setting: A Cost-Effectiveness Analysis. *Journal of Urology* 2018;200(6):1215-20. doi: https://dx.doi.org/10.1016/j.juro.2018.06.016
147. Dijkstra S, Govers TM, Hendriks RJ, et al. Cost-effectiveness of a new urinary biomarker-based risk score compared to standard of care in prostate cancer diagnostics - a decision analytical model. *BJU International* 2017;120(5):659-65. doi: https://dx.doi.org/10.1111/bju.13861
148. Heijnsdijk EAM, Denham D, de Koning HJ. The Cost-Effectiveness of Prostate Cancer Detection with the Use of Prostate Health Index. *Value in Health* 2016;19(2):153-57. doi: https://doi.org/10.1016/j.jval.2015.12.002
149. Nichol MB, Wu J, Huang J, et al. Cost-effectiveness of Prostate Health Index for prostate cancer detection. 2012;110(3):353-62. doi: 10.1111/j.1464-410X.2011.10751.x
150. Barnett CL, Davenport MS, Montgomery JS, et al. (18)F-Choline PET/mpMRI for Detection of Clinically Significant Prostate Cancer: Part 2. Cost-Effectiveness Analysis. *J Nucl Med* 2019;60(12):1705-12. doi: 10.2967/jnumed.119.225771 [published Online First: 2019/07/28]
151. Barnett CL, Davenport MS, Montgomery JS, et al. Cost-effectiveness of magnetic resonance imaging and targeted fusion biopsy for early detection of prostate cancer. *BJU International* 2018;122(1):50-58. doi: https://dx.doi.org/10.1111/bju.14151
152. Faria R, Soares MO, Spackman E, et al. Optimising the Diagnosis of Prostate Cancer in the Era of Multiparametric Magnetic Resonance Imaging: A Cost-effectiveness Analysis Based on the Prostate MR Imaging Study (PROMIS). *European Urology* 2018;73(1):23-30. doi: https://dx.doi.org/10.1016/j.eururo.2017.08.018
153. Pahwa S, Schiltz NK, Ponsky LE, et al. Cost-effectiveness of MR Imaging-guided Strategies for Detection of Prostate Cancer in Biopsy-Naive Men. *Radiology* 2017;285(1):157-66. doi: https://dx.doi.org/10.1148/radiol.2017162181
154. Venderink W, Govers TM, de Rooij M, et al. Cost-Effectiveness Comparison of Imaging-Guided Prostate Biopsy Techniques: Systematic Transrectal Ultrasound, Direct In-Bore MRI, and Image Fusion. *American Journal of Roentgenology* 2017;208(5):1058-63. doi: 10.2214/AJR.16.17322
155. Cerantola Y, Dragomir A, Tanguay S, et al. Cost-effectiveness of multiparametric magnetic resonance imaging and targeted biopsy in diagnosing prostate cancer. *Urologic Oncology: Seminars and Original Investigations* 2016;34(3):119.e1-19.e9. doi: https://doi.org/10.1016/j.urolonc.2015.09.010
156. de Rooij M, Crienen S, Witjes JA, et al. Cost-effectiveness of magnetic resonance (MR) imaging and MR-guided targeted biopsy versus systematic transrectal ultrasound-guided biopsy in diagnosing prostate cancer: a modelling study from a health care perspective. *Eur Urol* 2014;66:430.

157. NICE. Prostate cancer: diagnosis and management 2021 [Available from: https://www.nice.org.uk/guidance/ng131/update/NG131/documents/economic-report.
158. Nicholson A, Mahon J, Boland A, et al. The clinical effectiveness and cost-effectiveness of the PROGENSA (R) prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. 2015;19(87):1-191.
159. Mowatt G, Scotland G, Boachie C. The diagnostic accuracy and cost-effectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: a systematic review and economic evaluation. *Health Technol Assess* 2013;17:1.
160. Lesko LJ, Atkinson Jr AJArop, toxicology. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. 2001;41(1):347-66.
161. Greene KL, Albertsen PC, Babaian RJ, et al. Prostate Specific Antigen Best Practice Statement: 2009 Update. *Journal of Urology* 2009;182(5):2232-41. doi: doi:10.1016/j.juro.2009.07.093
162. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. 2013;190(2):419-26.
163. Mowatt G, Scotland G, Boachie C, et al. The diagnostic accuracy and cost-effectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(20):vii-xix, 1-281. doi: 10.3310/hta17200 [published Online First: 2013/05/24]
164. Van Neste L, Hendriks RJ, Dijkstra S, et al. Detection of high-grade prostate cancer using a urinary molecular biomarker–based risk score. *European urology* 2016;70(5):740-48.
165. Bouttell J, Teoh J, Chiu PK, et al. Economic evaluation of the introduction of the Prostate Health Index as a rule-out test to avoid unnecessary biopsies in men with prostate specific antigen levels of 4-10 in Hong Kong. *PloS one* 2019;14(4):e0215279-e79. doi: 10.1371/journal.pone.0215279
166. Teoh JY, Leung CH, Wang MH, et al. The cost-effectiveness of prostate health index for prostate cancer detection in Chinese men. *Prostate Cancer Prostatic Dis* 2020;23(4):615-21. doi: 10.1038/s41391-020-0243-1 [published Online First: 2020/07/02]
167. Eifler J, Alvarez J, Koyama T, et al. More judicious use of expectant management for localized prostate cancer during the last 2 decades. 2017;197(3 Part 1):614-20.
168. Liu J, Womble PR, Merdan S, et al. Factors influencing selection of active surveillance for localized prostate cancer. 2015;86(5):901-05.
169. Cooperberg MR, Broering JM, Carroll PRJJoCO. Time trends and local variation in primary treatment of localized prostate cancer. 2010;28(7):1117.
170. Tewari A, Gamito EJ, Crawford ED, et al. Biochemical recurrence and survival prediction models for the management of clinically localized prostate cancer. 2004;2(4):220-27.
171. Liu J, Womble PR, Merdan S, et al. Factors Influencing Selection of Active Surveillance for Localized Prostate Cancer. *Urology* 2015;86(5):901-5. doi: 10.1016/j.urology.2015.08.024 [published Online First: 2015/09/12]

172. Miñana B, Rodríguez-Antolín A, Gómez-Veiga F, et al. Treatment trends for clinically localised prostate cancer. National population analysis: GESCAP group. *Actas Urol Esp* 2016;40(4):209-16. doi: 10.1016/j.acuro.2015.09.005 [published Online First: 2016/01/03]
173. Trama A, Botta L, Nicolai N, et al. Prostate cancer changes in clinical presentation and treatments in two decades: an Italian population-based study. *Eur J Cancer* 2016;67:91-98. doi: 10.1016/j.ejca.2016.07.021 [published Online First: 2016/09/14]
174. Weissbach L, Stuerzebecher S, Mumperow E, et al. HAROW: the first comprehensive prospective observational study comparing treatment options in localized prostate cancer. *World J Urol* 2016;34(5):641-7. doi: 10.1007/s00345-015-1675-4 [published Online First: 2015/09/17]
175. Cussenot O, Rozet F, Ruffion A, et al. Prise en charge du cancer de la prostate: analyse rétrospective de 808 hommes biopsiés en France. *Progrès en urologie* 2013;23(5):347-55.
176. Léon P, Cancel-Tassin G, Koutlidis N, et al. Prévalence et diversité de la prise en charge des patients atteints de cancer de la prostate classés à faible risque selon la classification de d'Amico ou le score de CAPRA : étude française multicentrique. *Progrès en Urologie* 2017;27(3):158-65. doi: https://doi.org/10.1016/j.purol.2017.01.003
177. Gnanapragasam VJ, Lophatananon A, Wright KA, et al. Improving clinical risk stratification at diagnosis in primary prostate cancer: a prognostic modelling study. 2016;13(8):e1002063.
178. Postma R, Schröder FH, van Leenders GJLH, et al. Cancer Detection and Cancer Characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC) – Section Rotterdam: A Comparison of Two Rounds of Screening. *European Urology* 2007;52(1):89-97. doi: https://doi.org/10.1016/j.eururo.2007.01.030
179. Grey ADR, Chana MS, Popert R, et al. Diagnostic accuracy of magnetic resonance imaging (MRI) prostate imaging reporting and data system (PI-RADS) scoring in a transperineal prostate biopsy setting. *BJU International* 2015;115(5):728-35. doi: 10.1111/bju.12862
180. Siddiqui MM, Rais-Bahrami S, Turkbey B. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. *JAMA* 2015;313:390.
181. Stewart ST, Lenert L, Bhatnagar V, et al. Utilities for prostate cancer health states in men aged 60 and older. *Medical care* 2005:347-55.
182. Krahn M, Ritvo P, Irvine J, et al. Patient and Community Preferences for Outcomes in Prostate Cancer: Implications for Clinical Policy. *Medical Care* 2003;41(1):153-64.
183. Krahn MD, Bremner KE, Tomlinson G, et al. Utility and health-related quality of life in prostate cancer patients 12 months after radical prostatectomy or radiation therapy. *Prostate Cancer and Prostatic Diseases* 2009;12(4):361-68. doi: 10.1038/pcan.2009.32
184. Volk RJ, Cantor SB, Cass AR, et al. Preferences of husbands and wives for outcomes of prostate cancer screening and treatment. *Journal of general internal medicine* 2004;19(4):339-48. doi: 10.1111/j.1525-1497.2004.30046.x
185. Shimizu F, Fujino K, Ito YM, et al. Factors Associated with Variation in Utility Scores among Patients with Prostate Cancer. *Value in Health* 2008;11(7):1190-93. doi: https://doi.org/10.1111/j.1524-4733.2008.00336.x

186. Essink-Bot M-L, de Koning HJ, Nijs HG, et al. Short-term effects of population-based screening for prostate cancer on health-related quality of life. *Journal of National Cancer Institute* 1998;90(12):925-31.
187. Torvinen S, Färkkilä N, Sintonen H, et al. Health-related quality of life in prostate cancer. *Acta Oncologica* 2013;52(6):1094-101. doi: 10.3109/0284186X.2012.760848
188. Donovan JL, Hamdy FC, Lane JA, et al. Patient-Reported Outcomes after Monitoring, Surgery, or Radiotherapy for Prostate Cancer. *New England Journal of Medicine* 2016;375(15):1425-37. doi: 10.1056/NEJMoa1606221
189. Ware J, Snow KK, Kosinski M, et al. SF-36 health survey. *Manual and interpretation guide Boston: The Health Institute, New England Medical Center* 1993:10.6-10.7.
190. Heijnsdijk EA, Wever EM, Auvinen A. Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med* 2012;367:595.
191. De Haes J, de Koning HJ, van Oortmarssen GJ, et al. The impact of a breast cancer screening programme on quality-adjusted life-years. *International journal of cancer* 1991;49(4):538-44.
192. Korfage IJ, de Koning HJ, Roobol M, et al. Prostate cancer diagnosis: the impact on patients' mental health. *European Journal of Cancer* 2006;42(2):165-70.
193. Konski A, Sherman E, Krahn M, et al. Economic analysis of a phase III clinical trial evaluating the addition of total androgen suppression to radiation versus radiation alone for locally advanced prostate cancer (Radiation Therapy Oncology Group protocol 86-10). *International Journal of Radiation Oncology* Biology* Physics* 2005;63(3):788-94.
194. Calvert NW, Morgan AB, Catto JWF, et al. Effectiveness and cost-effectiveness of prognostic markers in prostate cancer. *British Journal of Cancer* 2003;88(1):31-35. doi: 10.1038/sj.bjc.6600630
195. Bennett CL, Matchar D, McCrory D, et al. Cost-effective models for flutamide for prostate carcinoma patients: Are they helpful to policy makers? 1996;77(9):1854-61. doi: 10.1002/(sici)1097-0142(19960501)77:9<1854::Aid-cncr15>3.0.Co;2-z
196. Zeliadt SB, Etzioni RD, Penson DF, et al. Lifetime implications and cost-effectiveness of using finasteride to prevent prostate cancer. *The American Journal of Medicine* 2005;118(8):850-57. doi: https://doi.org/10.1016/j.amjmed.2005.03.001
197. Cooperberg MR, Carroll PR, Klotz L. Active Surveillance for Prostate Cancer: Progress and Promise. 2011;29(27):3669-76. doi: 10.1200/jco.2011.34.9738
198. Konski A, Watkins-Bruner D, Brereton H, et al. Long-term hormone therapy and radiation is cost-effective for patients with locally advanced prostate carcinoma. 2006;106(1):51-57. doi: 10.1002/cncr.21575
199. Moeremans K, Caekelbergh K, Annemans L. Cost-Effectiveness Analysis of Bicalutamide (Casodex™)* *Casodex™ is a registered trademark for AstraZeneca. for Adjuvant Treatment of Early Prostate Cancer. *Value in Health* 2004;7(4):472-81. doi: https://doi.org/10.1111/j.1524-4733.2004.74010.x
200. Penson DF, RAMSEY S, VEENSTRA D, et al. The cost-effectiveness of combined androgen blockade with bicalutamide and luteinizing hormone releasing hormone agonist in men with metstatic prostate cancer. 2005;174(2):547-52. doi: doi:10.1097/01.ju.0000165569.48372.4c
201. Ramsey S, Veenstra D, Clarke L, et al. Is combined androgen blockade with bicalutamide cost-effective compared with combined androgen blockade with

flutamide? *Urology* 2005;66(4):835-39. doi: https://doi.org/10.1016/j.urology.2005.04.028
202. Sanghera S, Coast J, Martin RM, et al. Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models. *BMC Cancer* 2018;18(1):84. doi: https://dx.doi.org/10.1186/s12885-017-3974-1
203. Ara R, Brazier J, Peasgood T, et al. The Identification, Review and Synthesis of Health State Utility Values from the Literature. *PharmacoEconomics* 2017;35(1):43-55. doi: 10.1007/s40273-017-0547-8
204. Li C-k, Tong BCY, You JHS. Cost-effectiveness of culture-guided antimicrobial prophylaxis for the prevention of infections after prostate biopsy. *International Journal of Infectious Diseases* 2016;43:7-12. doi: https://doi.org/10.1016/j.ijid.2015.12.005
205. Cooperberg MR, Ramakrishna NR, Duff SB, et al. Primary treatments for clinically localised prostate cancer: a comprehensive lifetime cost-utility analysis. *BJU International* 2013;111(3):437-50. doi: 10.1111/j.1464-410X.2012.11597.x
206. Brown LC, Ahmed HU, Faria R, et al. Multiparametric MRI to improve detection of prostate cancer compared with transrectal ultrasound-guided prostate biopsy alone: the PROMIS study. *Health Technology Assessment (Winchester, England)* 2018;22(39):1-176. doi: https://dx.doi.org/10.3310/hta22390
207. Krahn MD, Mahoney JE, Eckman MH, et al. Screening for Prostate Cancer: A Decision Analytic View. *JAMA* 1994;272(10):773-80. doi: 10.1001/jama.1994.03520100035030
208. Bremner KE, Chong CAKY, Tomlinson G, et al. A Review and Meta-Analysis of Prostate Cancer Utilities. *Medical Decision Making* 2007;27(3):288-98. doi: 10.1177/0272989X07300604
209. Stewart ST, Lenert L, Bhatnagar V, et al. Utilities For Prostate Cancer Health States in Men Aged 60 and Older. *Medical Care* 2005;43(4):347-55. doi: 10.1097/01.mlr.0000156862.33341.45
210. Nichol MB, Wu J, Huang J, et al. Cost-effectiveness of Prostate Health Index for prostate cancer detection. *BJU International* 2012;110(3):353-62.
211. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical Prostatectomy or Watchful Waiting in Early Prostate Cancer. 2014;370(10):932-42. doi: 10.1056/NEJMoa1311593
212. Ades AE, Lu G, Claxton K. Expected Value of Sample Information Calculations in Medical Decision Modeling. *Medical Decision Making* 2004;24(2):207-27. doi: 10.1177/0272989X04263162
213. Bill-Axelson A, Holmberg L, Garmo H. Radical prostatectomy or watchful waiting in early prostate cancer. *N Engl J Med* 2014;370:932.
214. Draisma G, Boer R, Otto SJ, et al. Lead Times and Overdetection Due to Prostate-Specific Antigen Screening: Estimates From the European Randomized Study of Screening for Prostate Cancer. *JNCI: Journal of the National Cancer Institute* 2003;95(12):868-78. doi: 10.1093/jnci/95.12.868 %J JNCI: Journal of the National Cancer Institute
215. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. *New England Journal of Medicine* 2012;367(3):203-13.
216. James ND, Pirrie SJ, Pope AM, et al. Clinical Outcomes and Survival Following Treatment of Metastatic Castrate-Refractory Prostate Cancer With Docetaxel Alone or With Strontium-89, Zoledronic Acid, or Both: The TRAPEZE Randomized Clinical

Trial. *JAMA Oncology* 2016;2(4):493-99. doi: 10.1001/jamaoncol.2015.5570 %J JAMA Oncology
217. Nordström T, Aly M, Clements MS, et al. Prostate-specific antigen (PSA) testing is prevalent and increasing in Stockholm County, Sweden, despite no recommendations for PSA screening: results from a population-based study, 2003–2011. *European urology* 2013;63(3):419-25.
218. Van Hemelrijck M, Wigertz A, Sandin F, et al. Cohort profile: the national prostate cancer register of Sweden and prostate cancer data base Sweden 2.0. *International journal of epidemiology* 2013;42(4):956-67.
219. Hugosson J, Aus G, Lilja H, et al. Results of a Randomized, Population-Based Study of Biennial Screening Using Serum Prostate-Specific Antigen Measurement to Detect Prostate Carcinoma. *Cancer* 2004;100(7):1397-405. doi: 10.1002/cncr.20126
220. Hayes JH, Ollendorf DA, Pearson SD, et al. Active surveillance compared with initial treatment for men with low-risk prostate cancer: a decision analysis. *JAMA* 2010;304(21):2373-80. doi: 10.1001/jama.2010.1720
221. Thompson IM, Goodman PJ, Tangen CM, et al. Long-Term Survival of Participants in the Prostate Cancer Prevention Trial. 2013;369(7):603-10. doi: 10.1056/NEJMoa1215932
222. Aizer AA, Yu JB, Colberg JW, et al. Radical prostatectomy vs. intensity-modulated radiation therapy in the management of localized prostate adenocarcinoma. *Radiotherapy and Oncology* 2009;93(2):185-91. doi: https://doi.org/10.1016/j.radonc.2009.09.001
223. Klotz L, Emberton M. Management of low risk prostate cancer—active surveillance and focal therapy. *Nature Reviews Clinical Oncology* 2014;11(6):324-34. doi: 10.1038/nrclinonc.2014.73
224. Nam RK, Cheung P, Herschorn S, et al. Incidence of complications other than urinary incontinence or erectile dysfunction after radical prostatectomy or radiotherapy for prostate cancer: a population-based cohort study. *The Lancet Oncology* 2014;15(2):223-31. doi: https://doi.org/10.1016/S1470-2045(13)70606-5
225. Crook JM, O'Callaghan CJ, Duncan G, et al. Intermittent androgen suppression for rising PSA level after radiotherapy. *New England Journal of Medicine* 2012;367(10):895-903. doi: 10.1056/NEJMoa1201546
226. Dragomir A, Dinea D, Vanhuyse M, et al. Drug costs in the management of metastatic castration-resistant prostate cancer in Canada. 2014;14(1):252.
227. Pashayan N, Duffy SW, Pharoah P, et al. Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. *Br J Cancer* 2009;100(7):1198-204. doi: 10.1038/sj.bjc.6604973 [published Online First: 2009/03/19]
228. Wever EM, Hugosson J, Heijnsdijk EAM, et al. To be screened or not to be screened Modeling the consequences of PSA screening for the individual. *British Journal of Cancer*;107(5):778-84.
229. CCEMG - EPPI-Centre Cost Converter [Available from: https://eppi.ioe.ac.uk/costconversion/default.aspxconversion/default.aspx.
230. Bilcke J, Beutels P. Generating, Presenting, and Interpreting Cost-Effectiveness Results in the Context of Uncertainty: A Tutorial for Deeper Knowledge and Better Practice. *Medical Decision Making* 2021;42(4):421-35. doi: 10.1177/0272989X211045070

231. Chilcott J, Hummel S, Mildred M. Option appraisal: screening for prostate cancer. *Report to the UK National Screening Committee Sheffield: University of Sheffield, School of Health and Related Research* 2010
232. Imamura T, Yasunaga H. Economic evaluation of prostate cancer screening with prostate-specific antigen. *International Journal of Urology* 2008;15(4):285-88.
233. Lao C, Brown C, Rouse P, et al. Economic evaluation of prostate cancer screening: a systematic review. *Future Oncology* 2015;11(3):467-77.
234. Updated guidance for GPs on PSA testing for prostate cancer.: Public Health England; 2016 [updated 29 Mar 2016. Available from: https://phescreening.blog.gov.uk/2016/03/29/updated-guidance-for-gps-on-psa-testing-for-prostate-cancer/.
235. Seibert TM, Fan CC, Wang Y, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* 2018;360:j5757. doi: 10.1136/bmj.j5757 %J BMJ
236. Eklund M, Grönberg H, Nordström T. The STHLM3 prostate cancer diagnostic study: calibration, clarification, and comments. *Nature reviews Clinical oncology* 2016;13(6):394.
237. Sculpher M, Fenwick E, Claxton KJP. Assessing Quality in Decision Analytic Cost-Effectiveness Models. *Pharmacoeconomics* 2000;17(5):461-77. doi: 10.2165/00019053-200017050-00005
238. Yang Y, Abel L, Buchanan J, et al. Use of Decision Modelling in Economic Evaluations of Diagnostic Tests: An Appraisal and Review of Health Technology Assessments in the UK. *PharmacoEconomics - open* 2019;3(3):281-91. doi: 10.1007/s41669-018-0109-9
239. O'Mahony JF, Naber SK, Normand C, et al. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. *Value in Health* 2015;18(8):1138-51. doi: 10.1016/j.jval.2015.09.2939
240. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. 2012;32(5):678-89.
241. Tappenden P. Conceptual modelling for health economic model development. *HEDS Discussion Paper* 2012;12/05
242. Keeney E, Thom H, Turner E, et al. Using a Modified Delphi Approach to Gain Consensus on Relevant Comparators in a Cost-Effectiveness Model: Application to Prostate Cancer Screening. *PharmacoEconomics* 2021;39(5):589-600. doi: 10.1007/s40273-021-01009-6
243. Dalkey NC. The Delphi method: An experimental study of group opinion: RAND CORP SANTA MONICA CALIF, 1969.
244. Landeta J. Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change* 2006;73(5):467-82. doi: https://doi.org/10.1016/j.techfore.2005.09.002
245. Avella JR. Delphi panels: Research design, procedures, advantages, and challenges. *International Journal of Doctoral Studies* 2016;11(1):305-21.
246. Consensus statements on PSA testing in asymptomatic men in the UK - information for health professionals: Prostate Cancer UK; 2015 [Available from: https://prostatecanceruk.org/about-us/projects-and-policies/consensus-on-psa-testing/psa-consensus-for-health-professionals.

247. Giri VN, Knudsen KE, Kelly WK, et al. Role of Genetic Testing for Inherited Prostate Cancer Risk: Philadelphia Prostate Cancer Consensus Conference 2017. 2018;36(4):414-24. doi: 10.1200/jco.2017.74.1173
248. Scheltema MJ, Tay KJ, Postema AW, et al. Utilization of multiparametric prostate magnetic resonance imaging in clinical practice and focal therapy: report from a Delphi consensus project. *World Journal of Urology* 2017;35(5):695-701. doi: 10.1007/s00345-016-1932-1
249. Ritchie J, Lewis J, Nicholls CM, et al. Qualitative research practice: A guide for social science students and researchers: sage 2013.
250. Squires H, Chilcott J, Akehurst R, et al. A framework for developing the structure of public health economic models. 2016;19(5):588-601.
251. Husbands S, Jowett S, Barton P, et al. How qualitative methods can be used to inform model development. *PharmacoEconomics* 2017;35(6):607-12.
252. Rotondi A, Gustafson D. Gazing into the oracle: the Delphi method and its application to social policy and public health. 1996
253. Patton MQ. Qualitative research and evaluation methods . Thousand Oakes: ca: sage, 2002.
254. Kuper A, Lingard L, Levinson W. Critically appraising qualitative research. *BMJ* 2008;337:a1035.
255. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: Application of bootstrap data expansion. *BMC Med Res Methodol* 2005;5:1-12.
256. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009;42(2):377-81. doi: https://doi.org/10.1016/j.jbi.2008.08.010
257. Thorn JC, Brookes ST, Ridyard C, et al. Core Items for a Standardized Resource Use Measure: Expert Delphi Consensus Survey. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2018;21(6):640-49. doi: 10.1016/j.jval.2017.06.011
258. Boulkedid R, Abdoul H, Loustau M, et al. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PloS one* 2011;6(6):e20476.
259. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of clinical epidemiology* 2014;67(4):401-09.
260. Lindell MK, Brandt CJ, Whitney DJ. A Revised Index of Interrater Agreement for Multi-Item Ratings of a Single Target. *Applied Psychological Measurement* 1999;23(2):127-35. doi: 10.1177/01466219922031257
261. O'Neill TA. An Overview of Interrater Agreement on Likert Scales for Researchers and Practitioners. *Frontiers in psychology* 2017;8:777-77. doi: 10.3389/fpsyg.2017.00777
262. Sullivan W, Payne K. The Appropriate Elicitation of Expert Opinion in Economic Models. *PharmacoEconomics* 2011;29(6):455-59. doi: 10.2165/11589220-000000000-00000
263. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. *PharmacoEconomics* 2016;34(11):1161-72. doi: 10.1007/s40273-016-0425-9
264. Karlsson A, Jauhiainen A, Gulati R, et al. A natural history model for planning prostate cancer testing: Calibration and validation using Swedish registry data. *PLoS ONE*

[Electronic Resource] 2019;14(2):e0211918. doi: https://dx.doi.org/10.1371/journal.pone.0211918
265. Gulati R, Inoue L, Katcher J, et al. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. *Biostatistics* 2010;11(4):707-19.
266. Karlsson A, Jauhiainen A, Gulati R, et al. A natural history model for planning prostate cancer testing: Calibration and validation using Swedish registry data. *PLoS ONE* 2019;14(2):e0211918.
267. Etzioni R, Gulati R, Falcon S, et al. Impact of PSA Screening on the Incidence of Advanced Stage Prostate Cancer in the United States: A Surveillance Modeling Approach. *Medical Decision Making* 2008;28(3):323-31. doi: 10.1177/0272989X07312719
268. Gulati R, Gore JL, Etzioni R. Comparative Effectiveness of Alternative Prostate-Specific Antigen–Based Prostate Cancer Screening Strategies. *Annals of Internal Medicine* 2013;158(3):145-53. doi: 10.7326/0003-4819-158-3-201302050-00003
269. Thompson IM, Goodman PJ, Tangen CM, et al. The Influence of Finasteride on the Development of Prostate Cancer. *New England Journal of Medicine* 2003;349(3):215-24. doi: 10.1056/NEJMoa030660
270. Etzioni RD, Howlader N, Shaw PA, et al. Long-term effects of finasteride on prostate specific antigen levels: results from the prostate cancer prevention trial. *The Journal of urology* 2005;174(3):877-81.
271. Penney KL, Stampfer MJ, Jahn JL, et al. Gleason grade progression is uncommon. *Cancer Res* 2013;73(16):5163-8. doi: 10.1158/0008-5472.Can-13-0427
272. Sheridan TB, Carter HB, Wang W, et al. Change in prostate cancer grade over time in men followed expectantly for stage T1c disease. *J Urol* 2008;179(3):901-4; discussion 04-5. doi: 10.1016/j.juro.2007.10.062 [published Online First: 20080122]
273. Hussain M, Goldman B, Tangen C, et al. Prostate-specific antigen progression predicts overall survival in patients with metastatic prostate cancer: data from Southwest Oncology Group Trials 9346 (Intergroup Study 0162) and 9916. *J Clin Oncol* 2009;27(15):2450-6. doi: 10.1200/jco.2008.19.9810 [published Online First: 2009/04/22]
274. Karnes RJ, MacKintosh FR, Morrell CH, et al. Prostate-Specific Antigen Trends Predict the Probability of Prostate Cancer in a Very Large U.S. Veterans Affairs Cohort. *Front Oncol* 2018;8:296. doi: 10.3389/fonc.2018.00296 [published Online First: 2018/08/22]
275. Descazeaud A, Peyromaure M, Salin A, et al. Predictive Factors for Progression in Patients with Clinical Stage T1a Prostate Cancer in the PSA Era. *European Urology* 2008;53(2):355-62. doi: https://doi.org/10.1016/j.eururo.2007.06.020
276. Get Data Out: Prostate: National Disease Registration Service; 2022 [Available from: https://www.cancerdata.nhs.uk/getdataout/prostate accessed 23/11/2022.
277. Davis S, Stevenson, M., Tappenden, P., Wailoo, A.J. NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation. 2014
278. Wever EM, Draisma G, Heijnsdijk EAM, et al. Prostate-Specific Antigen Screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer–Rotterdam. *JNCI: Journal of the National Cancer Institute* 2010;102(5):352-55. doi: 10.1093/jnci/djp533

279. Nyame YA, Gulati R, Heijnsdijk EAM, et al. The Impact of Intensifying Prostate Cancer Screening in Black Men: A Model-Based Analysis. *J Natl Cancer Inst* 2021;113(10):1336-42. doi: 10.1093/jnci/djab072 [published Online First: 2021/05/09]
280. Getaneh AM, Heijnsdijk EAM, de Koning HJ. The comparative effectiveness of mpMRI and MRI-guided biopsy vs regular biopsy in a population-based PSA testing: a modeling study. *Sci Rep* 2021;11(1):1801. doi: 10.1038/s41598-021-81459-2 [published Online First: 2021/01/21]
281. Tsodikov A, Gulati R, de Carvalho TM, et al. Is prostate cancer different in black men? Answers from 3 natural history models. *Cancer* 2017;123(12):2312-19. doi: 10.1002/cncr.30687 [published Online First: 2017/04/25]
282. Heijnsdijk EAM, Gulati R, Tsodikov A, et al. Lifetime Benefits and Harms of Prostate-Specific Antigen-Based Risk-Stratified Screening for Prostate Cancer. *J Natl Cancer Inst* 2020;112(10):1013-20. doi: 10.1093/jnci/djaa001 [published Online First: 2020/02/19]
283. Roth JA, Gulati R, Gore JL, et al. Economic Analysis of Prostate-Specific Antigen Screening and Selective Treatment Strategies. *JAMA Oncol* 2016;2(7):890-8. doi: 10.1001/jamaoncol.2015.6275
284. Hao S, Heintz E, Östensson E, et al. Cost-Effectiveness of the Stockholm3 Test and Magnetic Resonance Imaging in Prostate Cancer Screening: A Microsimulation Study. *Eur Urol* 2022;82(1):12-19. doi: 10.1016/j.eururo.2021.12.021 [published Online First: 2022/02/01]
285. Cancer Survival in England: adults diagnosed between 2013 and 2017 and followed up to 2018: Office of National Statistics; 2018 [
286. Young GJ, Harrison S, Turner EL, et al. Prostate-specific antigen (PSA) testing of men in UK general practice: a 10-year longitudinal cohort study. *BMJ open* 2017;7(10):e017729-e29. doi: 10.1136/bmjopen-2017-017729
287. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology* 2015;44(3):827-36. doi: 10.1093/ije/dyv098
288. Caul S, Broggio J. Cancer registration statistics, England: 2017. *Off Natl Stat* 2019:1-16.
289. Heijnsdijk EAM, Gulati R, Lange JM, et al. Evaluation of Prostate Cancer Screening Strategies in a Low-Resource, High-risk Population in the Bahamas: JAMA Health Forum. 2022 May 20;3(5):e221116. doi: 10.1001/jamahealthforum.2022.1116. eCollection 2022 May.
290. Powell MJ. The BOBYQA algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* 2009;26
291. Nash JC. On Best Practice Optimization Methods in R. *Journal of Statistical Software* 2014;60(2):1 - 14. doi: 10.18637/jss.v060.i02
292. Hendrix N, Gulati R, Jiao B, et al. Clarifying the Trade-Offs of Risk-Stratified Screening for Prostate Cancer: A Cost-Effectiveness Study. *American Journal of Epidemiology* 2021 doi: 10.1093/aje/kwab155
293. Callender T, Emberton M, Morris S, et al. Polygenic risk-tailored screening for prostate cancer: A benefit–harm and cost-effectiveness modelling study. *PLOS Medicine* 2019;16(12):e1002998. doi: 10.1371/journal.pmed.1002998

294. Pharoah PDP, Antoniou A, Bobrow M, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nature Genetics* 2002;31(1):33-36. doi: 10.1038/ng853
295. Hao S, Karlsson A, Heintz E, et al. Cost-Effectiveness of Magnetic Resonance Imaging in Prostate Cancer Screening: A Microsimulation Study. *Value in Health* 2021;24(12):1763-72. doi: 10.1016/j.jval.2021.06.001
296. Drost FH, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. *Cochrane Database Syst Rev* 2019;4(4):Cd012663. doi: 10.1002/14651858.CD012663.pub2 [published Online First: 2019/04/26]
297. Hao. Cost-effectiveness of magnetic resonance imaging in prostate cancer screening: a microsimulation study. *Value in Health* 2021;in press
298. National Cancer Registration and Analysis Service 2021 [Available from: https://www.cancerdata.nhs.uk/getdataout/prostate accessed 13/5/21.
299. Venderink W, Govers TM, De Rooij M, et al. Cost-effectiveness comparison of imaging-guided prostate biopsy techniques: Systematic transrectal ultrasound, direct in-bore MRI, and image fusion. *American Journal of Roentgenology* 2017;208(5):1058-63.
300. National life tables: UK: Office for National Statistics; 2020 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
301. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *New England Journal of Medicine* 2016;375(15):1415-24. doi: 10.1056/NEJMoa1606220
302. Dixon P, Keeney E, Taylor JC, et al. Can polygenic risk scores contribute to cost-effective cancer screening? A systematic review. *Genet Med* 2022 doi: 10.1016/j.gim.2022.04.020 [published Online First: 2022/05/17]
303. Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliat Med* 2015;29(10):899-907. doi: 10.1177/0269216315595203 [published Online First: 2015/07/23]
304. Hospital episode statistics, 2012–2013 London: Department of Health; 2014 [Available from: https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/hospital-episode-statistics-admitted-patient-care-england-2012-13.
305. NICE. Developing NICE guidelines: the manual 2022 [Available from: https://www.nice.org.uk/process/pmg20/chapter/introductionPublished: 31 October 2014 Last updated: 18 January 2022.
306. Li L, Severens JL, Mandrik O. Disutility associated with cancer screening programs: A systematic review. *PLOS ONE* 2019;14(7):e0220148. doi: 10.1371/journal.pone.0220148
307. Loeb S, Curnyn C, Walter D, et al. Health state utilities among contemporary prostate cancer patients on active surveillance. *Transl Androl Urol* 2018;7(2):197-202. doi: 10.21037/tau.2017.03.80 [published Online First: 2018/05/08]
308. Hall PS, Hamilton P, Hulme CT, et al. Costs of cancer care for use in economic evaluation: a UK analysis of patient-level routine health system data. *British Journal of Cancer* 2015;112(5):948-56. doi: 10.1038/bjc.2014.644

309. 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. *International Journal of Cancer* 2005;116(2):291-96.
310. Glazener C, Boachie C, Buckley B, et al. Urinary incontinence in men after formal one-to-one pelvic-floor muscle training following radical prostatectomy or transurethral resection of the prostate (MAPS): two parallel randomised controlled trials. *Lancet* 2011;378(9788):328-37. doi: 10.1016/s0140-6736(11)60751-4 [published Online First: 2011/07/12]
311. Watson E, Shinkins B, Frith E, et al. Symptoms, unmet needs, psychological well-being and health status in survivors of prostate cancer: implications for redesigning follow-up. *BJU International* 2016;117(6B):E10-E19. doi: https://doi.org/10.1111/bju.13122
312. Magnus A, Isaranuwatchai W, Mihalopoulos C, et al. A Systematic Review and Meta-Analysis of Prostate Cancer Utility Values of Patients and Partners Between 2007 and 2016. *MDM Policy Pract* 2019;4(1):2381468319852332. doi: 10.1177/2381468319852332 [published Online First: 2019/06/14]
313. Wu EQ, Mulani P, Farrell MH, et al. Mapping FACT-P and EORTC QLQ-C30 to patient health status measured by EQ-5D in metastatic hormone-refractory prostate cancer patients. *Value Health* 2007;10(5):408-14. doi: 10.1111/j.1524-4733.2007.00195.x [published Online First: 2007/09/25]
314. Janssen B, Szende A. Population Norms for the EQ-5D. In: Szende A, Janssen B, Cabases J, eds. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht: Springer Netherlands 2014:19-30.
315. Dolan P, Gudex C, Kind P, et al. A social tariff for EuroQol: results from a UK general population survey, 1995.
316. 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(2):291-6. doi: 10.1002/ijc.21043 [published Online First: 2005/04/01]
317. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35(11):1095-108. doi: 10.1097/00005650-199711000-00002 [published Online First: 1997/11/21]
318. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health* 2012;15(5):708-15. doi: 10.1016/j.jval.2012.02.008 [published Online First: 2012/08/08]
319. Färkkilä N, Torvinen S, Roine RP, et al. Health-related quality of life among breast, prostate, and colorectal cancer patients with end-stage disease. *Qual Life Res* 2014;23(4):1387-94. doi: 10.1007/s11136-013-0562-y [published Online First: 2013/11/02]
320. EuroQol-a new facility for the measurement of health-related quality of life: EuroQol Group; 1990 [199-208]. accessed 3 16.
321. Penson David F, Ramsey S, Veenstra D, et al. THE COST-EFFECTIVENESS OF COMBINED ANDROGEN BLOCKADE WITH BICALUTAMIDE AND LUTEINIZING HORMONE RELEASING HORMONE AGONIST IN MEN WITH METASTATIC PROSTATE CANCER. *Journal of Urology* 2005;174(2):547-52. doi: 10.1097/01.ju.0000165569.48372.4c
322. Bayoumi AM, Brown AD, Garber AM. Cost-effectiveness of androgen suppression therapies in advanced prostate cancer. *Journal of the National Cancer Institute* 2000;92(21):1731-39.

323. Kind P, Dolan P, Gudex C, et al. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998;316(7133):736-41. doi: 10.1136/bmj.316.7133.736
324. Wilt TJ, Brawer MK, Jones KM, et al. Radical Prostatectomy versus Observation for Localized Prostate Cancer. 2012;367(3):203-13. doi: 10.1056/NEJMoa1113162
325. James ND, Spears MR, Clarke NW, et al. Survival with Newly Diagnosed Metastatic Prostate Cancer in the "Docetaxel Era": Data from 917 Patients in the Control Arm of the STAMPEDE Trial (MRC PR08, CRUK/06/019). *Eur Urol* 2015;67(6):1028-38. doi: 10.1016/j.eururo.2014.09.032 [published Online First: 2014/10/11]
326. Gnanapragasam VJ, Lophatananon A, Wright KA, et al. Improving Clinical Risk Stratification at Diagnosis in Primary Prostate Cancer: A Prognostic Modelling Study. *PLoS medicine* 2016;13(8):e1002063-e63. doi: 10.1371/journal.pmed.1002063
327. Lasdon LS, Fox RL, Ratner MW. Nonlinear optimization using the generalized reduced gradient method. *Revue française d'automatique, informatique, recherche opérationnelle Recherche opérationnelle* 1974;8(V3):73-103.
328. Loeb S, Zhou Q, Siebert U, et al. Active Surveillance Versus Watchful Waiting for Localized Prostate Cancer: A Model to Inform Decisions. *European Urology* 2017;72(6):899-907. doi: https://doi.org/10.1016/j.eururo.2017.07.018
329. Hamid S, Donaldson IA, Hu Y, et al. The SmartTarget Biopsy Trial: A prospective, within-person randomised, blinded trial comparing the accuracy of visual-registration and magnetic resonance imaging/ultrasound image-fusion targeted biopsies for trostate cancer risk stratification. *Eur Urol* 2019;75(5):733-40. doi: 10.1016/j.eururo.2018.08.007 [published Online First: 2018/12/12]
330. Murthy V, Sonni I, Jariwala N, et al. The Role of PSMA PET/CT and PET/MRI in the Initial Staging of Prostate Cancer. *Eur Urol Focus* 2021;7(2):258-66. doi: 10.1016/j.euf.2021.01.016 [published Online First: 2021/02/06]
331. Hofman MS, Lawrentschuk N, Francis RJ, et al. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *The Lancet* 2020;395(10231):1208-16. doi: 10.1016/s0140-6736(20)30314-7
332. Claxton K. Exploring Uncertainty in Cost-Effectiveness Analysis. *PharmacoEconomics* 2008;26(9):781-98. doi: 10.2165/00019053-200826090-00008
333. Keeney E, Sanghera S, Martin RM, et al. Cost-Effectiveness Analysis of Prostate Cancer Screening in the UK: A Decision Model Analysis Based on the CAP Trial. *PharmacoEconomics* 2022 doi: 10.1007/s40273-022-01191-1
334. Pataky R, Gulati R, Etzioni R, et al. Is prostate cancer screening cost-effective? A microsimulation model of prostate-specific antigen-based screening for British Columbia, Canada (Provisional abstract). 2014; 135(4). http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-22014026830/frame.html http://onlinelibrary.wiley.com/store/10.1002/ijc.28732/asset/ijc28732.pdf?v=1&t=infwixm5&s=68e3a8b5fc4ae1bfdda9dc4917b0d77d5e8e99b3 (accessed 2014).
335. Roth JA, Gulati R, Gore JL, et al. Economic analysis of prostate-specific antigen screening and selective treatment strategies. *JAMA Oncol* 2016;2:890.
336. Auvinen A, Rannikko A, Taari K, et al. A randomized trial of early detection of clinically significant prostate cancer (ProScreen): study design and rationale. 2017;32(6):521-27. doi: 10.1007/s10654-017-0292-5

337. Purysko AS, Rosenkrantz AB, Barentsz JO, et al. PI-RADS Version 2: A Pictorial Update. *RadioGraphics* 2016;36(5):1354-72. doi: 10.1148/rg.2016150234
338. Hoffman RM, Denberg T, Hunt WC, et al. Prostate cancer testing following a negative prostate biopsy: over testing the elderly. *J Gen Intern Med* 2007;22(8):1139-43. doi: 10.1007/s11606-007-0248-7 [published Online First: 2007/06/08]
339. Busby JE, Evans CP. Determining variables for repeat prostate biopsy. *Prostate Cancer and Prostatic Diseases* 2004;7(2):93-98. doi: 10.1038/sj.pcan.4500708
340. Essid MA, Chakroun M, Nouhaud FX, et al. Evolution of prostate cancer diagnosis: retrospective analysis of magnetic resonance imaging/ultrasound fusion guided biopsies protocol in routine practice and patients management. *Transl Androl Urol* 2020;9(2):629-36.
341. Krahn M, Bremner KE, Tomlinson G, et al. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Quality of Life Research* 2006;16(3):509. doi: 10.1007/s11136-006-9132-x
342. Krahn M, Ritvo P, Irvine J, et al. Construction of the Patient-Oriented Prostate Utility Scale (PORPUS): a multiattribute health state classification system for prostate cancer. *Journal of clinical epidemiology* 2000;53(9):920-30.
343. Tomlinson G, Bremner KE, Ritvo P, et al. Development and Validation of a Utility Weighting Function for the Patient-Oriented Prostate Utility Scale (PORPUS). *Medical Decision Making* 2012;32(1):11-30. doi: 10.1177/0272989x11407203
344. Ku J, Krahn M, Trachtenberg J, et al. Changes in health utilities and health-related quality of life over 12 months following radical prostatectomy. *Canadian Urological Association Journal* 2009;3(6):445-52.
345. Avila M, Pardo Y, Castells M, et al. Adaptation and validation of the Spanish version of the Patient-Oriented Prostate Utility Scale (PORPUS). *Qual Life Res* 2014;23(9):2481-7. doi: 10.1007/s11136-014-0701-0 [published Online First: 2014/05/03]
346. Bremner KE, Mitsakakis N, Wilson L, et al. Predicting utility scores for prostate cancer: mapping the Prostate Cancer Index to the Patient-Oriented Prostate Utility Scale (PORPUS). *Prostate Cancer Prostatic Dis* 2014;17(1):47-56. doi: 10.1038/pcan.2013.44 [published Online First: 2013/10/16]
347. Krahn MD, Bremner KE, Alibhai SM, et al. A reference set of health utilities for long-term survivors of prostate cancer: population-based data from Ontario, Canada. *Qual Life Res* 2013;22(10):2951-62. doi: 10.1007/s11136-013-0401-1 [published Online First: 2013/04/09]
348. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ* 1996;5(6):513-24. doi: 10.1002/(sici)1099-1050(199611)5:6<513::Aid-hec237>3.0.Co;2-9 [published Online First: 1996/11/01]
349. Accelerated Value of Information Release version 2.0.8: University of Sheffield; 2015 [Available from: https://savi.shef.ac.uk/SAVI/.
350. Rasmussen CE. Gaussian processes in machine learning: Springer; 2003 [63-71]. Available from: https://link.springer.com/chapter/10.1007/978-3-540-28650-9_4.
351. Alarid-Escudero F, Enns EA, Kuntz KM, et al. "Time Traveling Is Just Too Dangerous" but Some Methods Are Worth Revisiting: The Advantages of Expected Loss Curves Over Cost-Effectiveness Acceptability Curves and Frontier. *Value Health* 2019;22(5):611-18. doi: 10.1016/j.jval.2019.02.008 [published Online First: 2019/05/21]

352. Heijnsdijk EAM, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. *Journal of the National Cancer Institute* 2014;107(1):366-66. doi: 10.1093/jnci/dju366
353. Hummel S, Chilcott J. Option Appraisal: Screening for prostate cancer model update. *Sheffield: In: ScHARR, University of Sheffield* 2013
354. Roobol MJ, Kranse R, Bangma CH, et al. Screening for prostate cancer: results of the Rotterdam section of the European randomized study of screening for prostate cancer. *Eur Urol* 2013;64(4):530-9. doi: 10.1016/j.eururo.2013.05.030 [published Online First: 2013/06/14]
355. Mahase E. Is the UK really ready to roll out prostate cancer screening? *BMJ* 2023;381:p1062. doi: 10.1136/bmj.p1062
356. Baltussen R, Adam T, Tan Torres T, et al. Generalized cost-effectiveness analysis: a guide. *Geneva: World Health Organization* 2002;71
357. Johns B, Baltussen R, Hutubessy R. Programme costs in the economic evaluation of health interventions. *Cost Effectiveness and Resource Allocation* 2003;1(1):1. doi: 10.1186/1478-7547-1-1
358. Langer A, Holle R, John J. Specific guidelines for assessing and improving the methodological quality of economic evaluations of newborn screening. *BMC Health Services Research* 2012;12(1):300. doi: 10.1186/1472-6963-12-300
359. Stark JR, Perner S, Stampfer MJ, et al. Gleason score and lethal prostate cancer: does 3 + 4 = 4 + 3? *J Clin Oncol* 2009;27(21):3459-64. doi: 10.1200/jco.2008.20.4669 [published Online First: 2009/05/13]
360. Wever EM, Hugosson J, Heijnsdijk EA, et al. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. *Br J Cancer* 2012;107(5):778-84. doi: 10.1038/bjc.2012.317 [published Online First: 2012/07/19]
361. Wever EM, Draisma G, Heijnsdijk EA, et al. How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening. *Medical Decision Making* 2011;31(4):550-8.
362. Squires H, Chilcott J, Akehurst R, et al. A Framework for Developing the Structure of Public Health Economic Models. *Value in Health* 2016;19(5):588-601. doi: https://doi.org/10.1016/j.jval.2016.02.011
363. Munn Z, Peters MDJ, Stern C, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology* 2018;18(1):143. doi: 10.1186/s12874-018-0611-x
364. Hatswell AJ, Chandler F. Sharing is Caring: The Case for Company-Level Collaboration in Pharmacoeconomic Modelling. *Pharmacoeconomics* 2017;35(8):755-57. doi: 10.1007/s40273-017-0516-2 [published Online First: 2017/05/22]
365. Sampson CJ, Arnold R, Bryan S, et al. Transparency in Decision Modelling: What, Why, Who and How? *Pharmacoeconomics* 2019;37(11):1355-69. doi: 10.1007/s40273-019-00819-z [published Online First: 2019/06/27]
366. Arnold RJG, Ekins S. Ahead of Our Time: Collaboration in Modeling Then and Now. *PharmacoEconomics* 2017;35(9):975-76. doi: 10.1007/s40273-017-0532-2
367. Prakash MK, Lang B, Heinrich H, et al. CMOST: an open-source framework for the microsimulation of colorectal cancer screening strategies. *BMC Medical Informatics and Decision Making* 2017;17(1):80. doi: 10.1186/s12911-017-0458-9
368. Gray E, Donten A, Karssemeijer N, et al. Evaluation of a Stratified National Breast Screening Program in the United Kingdom: An Early Model-Based Cost-Effectiveness

Analysis. *Value in Health* 2017;20(8):1100-09. doi: https://doi.org/10.1016/j.jval.2017.04.012
369. Cohen JT, Neumann PJ, Wong JB. A call for open-source cost-effectiveness analysis: American College of Physicians, 2017:432-33.
370. Dunlop WC, Mason N, Kenworthy J, et al. Benefits, challenges and potential strategies of open source health economic models. *Pharmacoeconomics* 2017;35(1):125-28.
371. Pouwels XGLV, Sampson CJ, Arnold RJG, et al. Opportunities and Barriers to the Development and Use of Open Source Health Economic Models: A Survey. *Value in Health* 2022;25(4):473-79. doi: 10.1016/j.jval.2021.10.001
372. McManus E, Turner D, Gray E, et al. Barriers and Facilitators to Model Replication Within Health Economics. *Value Health* 2019;22(9):1018-25. doi: 10.1016/j.jval.2019.04.1928 [published Online First: 2019/09/13]
373. Loeve F, Boer R, van Oortmarssen GJ, et al. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Comput Biomed Res* 1999;32(1):13-33. doi: 10.1006/cbmr.1998.1498 [published Online First: 1999/03/06]
374. Loeve F, Brown ML, Boer R, et al. Endoscopic colorectal cancer screening: a cost-saving analysis. *J Natl Cancer Inst* 2000;92(7):557-63. doi: 10.1093/jnci/92.7.557 [published Online First: 2000/04/06]
375. Bojke L, Soares MO, Claxton K, et al. Reference Case Methods for Expert Elicitation in Health Care Decision Making. *Medical Decision Making* 2021;42(2):182-93. doi: 10.1177/0272989X211028236
376. Thom H, Jackson C, Welton N, et al. Using Parameter Constraints to Choose State Structures in Cost-Effectiveness Modelling. *Pharmacoeconomics* 2017;35(9):951-62. doi: 10.1007/s40273-017-0501-9 [published Online First: 2017/03/28]
377. Ghabri S, Cleemput I, Josselin J-M. Towards a New Framework for Addressing Structural Uncertainty in Health Technology Assessment Guidelines. *PharmacoEconomics* 2018;36(2):127-30. doi: 10.1007/s40273-017-0603-4
378. Jackson CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc Ser A Stat Soc* 2009;172(2):383-404. doi: 10.1111/j.1467-985X.2008.00573.x [published Online First: 2009/04/22]
379. Nicholson A, Mahon J, Bol, et al. The clinical effectiveness and cost-effectiveness of the PROGENSA prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England)* 2015;19(87):i-xxxi, 1-191.
380. Wright SJ, Newman WG, Payne K. Quantifying the Impact of Capacity Constraints in Economic Evaluations: An Application in Precision Medicine. *Med Decis Making* 2022;42(4):538-53. doi: 10.1177/0272989x211053792 [published Online First: 2021/10/26]
381. Haji Ali Afzali H, Karnon J, Theou O, et al. Structuring a conceptual model for cost-effectiveness analysis of frailty interventions. *PLOS ONE* 2019;14(9):e0222049. doi: 10.1371/journal.pone.0222049
382. Catalona WJ, Partin AW, Sanda MG, et al. A multicenter study of [-2] pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. 2011;185(5):1650-55.

383. Dakna M, He Z, Yu WC, et al. Technical, bioinformatical and statistical aspects of liquid chromatography–mass spectrometry (LC–MS) and capillary electrophoresis-mass spectrometry (CE-MS) based clinical proteomics: A critical assessment. *Journal of Chromatography B* 2009;877(13):1250-58. doi: https://doi.org/10.1016/j.jchromb.2008.10.048
384. Konety B, Zappala SM, Parekh DJ, et al. The 4Kscore® Test Reduces Prostate Biopsy Rates in Community and Academic Urology Practices. *Reviews in urology* 2015;17(4):231-40.
385. McKiernan J, Donovan MJ, O'Neill V, et al. A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy. 2016;2(7):882-89.
386. Bussemakers MJG, van Bokhoven A, Verhaegh GW, et al. A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer. *Cancer Research* 1999;59(23):5975.
387. Løgager VB. MRI of the Prostate: A Practical Approach: SAGE Publications Sage UK: London, England, 2018.
388. Addicott B, Foster BR, Johnson C, et al. Direct magnetic resonance imaging-guided biopsy of the prostate: lessons learned in establishing a regional referral center. *Transl Androl Urol* 2017;6(3):395-405. doi: 10.21037/tau.2017.01.14
389. 18F-choline: National Cancer Institute; 2021 [Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/18f-choline accessed 11/4/21.
390. Chiu PK, Roobol MJ, Teoh JY, et al. Prostate health index (PHI) and prostate-specific antigen (PSA) predictive models for prostate cancer in the Chinese population and the role of digital rectal examination-estimated prostate volume. *International urology and nephrology* 2016;48(10):1631-37.
391. Lecornet E, Ahmed HU, Hu Y, et al. The Accuracy of Different Biopsy Strategies for the Detection of Clinically Important Prostate Cancer: A Computer Simulation. *Journal of Urology* 2012;188(3):974-80. doi: doi:10.1016/j.juro.2012.04.104
392. Catalona WJ, Partin AW, Sanda MG, et al. A multicenter study of [-2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. *The Journal of urology* 2011;185(5):1650-55. doi: 10.1016/j.juro.2010.12.032 [published Online First: 2011/03/17]
393. Parekh DJ, Punnen S, Sjoberg DD, et al. A Multi-institutional Prospective Trial in the USA Confirms that the 4Kscore Accurately Identifies Men with High-grade Prostate Cancer. *European Urology* 2015;68(3):464-70. doi: https://doi.org/10.1016/j.eururo.2014.10.021
394. McKiernan J, Donovan MJ, O'Neill V, et al. A Novel Urine Exosome Gene Expression Assay to Predict High-grade Prostate Cancer at Initial Biopsy. *JAMA Oncology* 2016;2(7):882-89. doi: 10.1001/jamaoncol.2016.0097 %J JAMA Oncology
395. Epstein JI, Feng Z, Trock BJ, et al. Upgrading and Downgrading of Prostate Cancer from Biopsy to Radical Prostatectomy: Incidence and Predictive Factors Using the Modified Gleason Grading System and Factoring in Tertiary Grades. *European Urology* 2012;61(5):1019-24. doi: https://doi.org/10.1016/j.eururo.2012.01.050
396. Iwazawa J MT, Sassa S, Ohue S. Prostate cancer detection with MRI: is dynamic contrast enhanced imaging necessary in addition to diffusion-weighted imaging? *Diagn Interv Radiol* 2011;17(3):243-48.

397. Tamada T, Sone T, Higashi H, et al. Prostate Cancer Detection in Patients With Total Serum Prostate-Specific Antigen Levels of 4–10 ng/mL: Diagnostic Efficacy of Diffusion-Weighted Imaging, Dynamic Contrast-Enhanced MRI, and T2-Weighted Imaging. *American Journal of Roentgenology* 2011;197(3):664-70. doi: 10.2214/AJR.10.5923
398. Schoots IG, Roobol MJ, Nieboer D, et al. Magnetic Resonance Imaging–targeted Biopsy May Enhance the Diagnostic Accuracy of Significant Prostate Cancer Detection Compared to Standard Transrectal Ultrasound-guided Biopsy: A Systematic Review and Meta-analysis. *European Urology* 2015;68(3):438-50. doi: https://doi.org/10.1016/j.eururo.2014.11.037
399. Mendhiratta N, Rosenkrantz AB, Meng X, et al. Magnetic Resonance Imaging-Ultrasound Fusion Targeted Prostate Biopsy in a Consecutive Cohort of Men with No Previous Biopsy: Reduction of Over Detection through Improved Risk Stratification. 2015;194(6):1601-06. doi: doi:10.1016/j.juro.2015.06.078
400. Pokorny MR, de Rooij M, Duncan E. Prospective study of diagnostic accuracy comparing prostate cancer detection by transrectal ultrasound-guided biopsy versus magnetic resonance (MR) imaging with subsequent MR-guided biopsy in men without previous prostate biopsies. *Eur Urol* 2014;66:22.
401. Abd-Alazeez M, Kirkham A, Ahmed HU. Performance of multiparametric MRI in men at risk of prostate cancer before the first biopsy: a paired validating cohort study using template prostate mapping biopsies as the reference standard. *Prostate Cancer Prostatic Dis* 2014;17:40.
402. Hoeks CMA, Schouten MG, Bomers JGR, et al. Three-Tesla Magnetic Resonance–Guided Prostate Biopsy in Men With Increased Prostate-Specific Antigen and Repeated, Negative, Random, Systematic, Transrectal Ultrasound Biopsies: Detection of Clinically Significant Prostate Cancers. *European Urology* 2012;62(5):902-09. doi: https://doi.org/10.1016/j.eururo.2012.01.047
403. Roethke M, Anastasiadis AG, Lichy M, et al. MRI-guided prostate biopsy detects clinically significant cancer: analysis of a cohort of 100 patients after previous negative TRUS biopsy. 2012;30(2):213-18. doi: 10.1007/s00345-011-0675-2
404. Franiel T, Stephan C, Erbersdobler A, et al. Areas Suspicious for Prostate Cancer: MR–guided Biopsy in Patients with at Least One Transrectal US-guided Biopsy with a Negative Finding—Multiparametric MR Imaging for Detection and Biopsy Planning. 2011;259(1):162-72. doi: 10.1148/radiol.10101251
405. Hambrock T, Somford DM, Hoeks C. Magnetic resonance imaging guided prostate biopsy in men with repeat negative biopsies and increased prostate specific antigen. *J Urol* 2010;183:520.
406. Taira AV, Merrick GS, Bennett A, et al. Transperineal Template-guided Mapping Biopsy as a Staging Procedure to Select Patients Best Suited for Active Surveillance. 2013;36(2):116-20. doi: 10.1097/COC.0b013e31823fe639
407. de Rooij M, Hamoen EHJ, Fütterer JJ, et al. Accuracy of Multiparametric MRI for Prostate Cancer Detection: A Meta-Analysis. *American Journal of Roentgenology* 2014;202(2):343-51. doi: 10.2214/AJR.13.11046
408. Scattoni V, Raber M, Capitanio U, et al. The Optimal Rebiopsy Prostatic Scheme Depends on Patient Clinical Characteristics: Results of a Recursive Partitioning Analysis Based on a 24-Core Systematic Scheme. *European Urology* 2011;60(4):834-41. doi: https://doi.org/10.1016/j.eururo.2011.07.036

409. Davenport MS, Montgomery JS, Kunju LP, et al. ¹⁸F-Choline PET/mpMRI for Detection of Clinically Significant Prostate Cancer: Part 1. Improved Risk Stratification for MRI-Guided Transrectal Prostate Biopsies. *Journal of Nuclear Medicine* 2020;61(3):337-43. doi: 10.2967/jnumed.119.225789
410. Haas GP, Delongchamps NB, Jones RF, et al. Needle Biopsies on Autopsy Prostates: Sensitivity of Cancer Detection Based on True Prevalence. *JNCI: Journal of the National Cancer Institute* 2007;99(19):1484-89. doi: 10.1093/jnci/djm153 %J JNCI: Journal of the National Cancer Institute
411. Teoh JYC, Chiu PKF, Yip SY, et al. The cost-effectiveness of prostate health index for prostate cancer detection in chinese men. *International Journal of Urology* 2018;25:342.
412. Schiffer E, Bick C, Grizelj B, et al. Urinary proteome analysis for prostate cancer diagnosis: cost-effective application in routine clinical practice in Germany. *International Journal of Urology* 2012;19(2):118-25.
413. Mowatt G, Scotl, G, et al. The diagnostic accuracy and costeffectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: A systematic review and economic evaluation. *Health Technology Assessment* 2013;17(20):i-ix+281.

Appendix 1: Descriptions of biomarkers and biopsy methods

Table 20. Descriptions of recently developed biomarkers

Biomarker	Definition
Prostate Health Index (PHI)	Blood biomarker calculated by a test analyser from the combination of total PSA (tPSA), free PSA (fPSA), and [-2]proPSA assays. It is used to calculate the probability of PCa and as an aid in distinguishing PCa from benign prostatic conditions for men with a borderline PSA test (e.g. PSA 2-10 ng/mL or 4-10 ng/mL) and non-suspicious digital rectal examination.[382]
SelectMDx (SelectMDx; MDxHealth, Inc., Irvine, CA, USA)	Urinary molecular biomarker-based risk score developed to identify patients that are at risk of harbouring high-grade PCa (Gleason score ≥7).[164] This risk score is based on the urinary homeobox C6 (HOXC6) and distal-less homeobox 1 (DLX1) mRNA signature in combination with serum PSA level, PSA density, and other clinical risk factors such as age, prior cancer-negative biopsies, DRE, and family history
Urinary Proteome Analysis for PCa diagnosis (UPA-PC)	Based on capillary electrophoresis mass spectrometry (CE/MS) and allows proteome analyses of prostatic secretions in first stream urine (first 10–15 mL urine fraction) to distinguish patients with positive PSA and/or DRE with PCa from those without PCa. The test simultaneously determines 12 biomarkers combined as a PC-specific multi-biomarker profile.[383]

4Kscore® Test (OPKO Diagnostics, LLC)	Incorporates measured blood levels of four kallikrein proteins: total PSA, free PSA, intact PSA, and human kallikrein 2 plus clinical information (age, DRE findings, and a history of prior negative biopsy result) into an algorithm to calculate an individual man's percentage risk (< 1% to > 95%) of having Gleason score ≥ 7 if a prostate biopsy were to be performed.[384]
ExoDx Prostate (IntelliScore) (EPI)	Urine exosome gene expression test, which utilizes exosomal RNA expression levels of three genes to predict the likelihood of having high-grade PCa of Grade Group 2 or greater.[385]
PCA3	A segment of noncoding messenger ribonucleic acid (mRNA) from chromosome 9q21–22 that is overexpressed by more than 95% of all PCas tested [386].
PSA density	Serum PSA level (ng/mL) divided by volume of the prostate gland (mL)[387]
Stockholm3	Blood-based prostate cancer test that analyzes Total PSA, free PSA, HK2, MSMB and MIC1, more than 100 genetic markers as well as age, earlier prostate biopsy, family history of prostate biopsy and use of 5-alfareducase inhibitors. In addition digital rectal examination (DRE) and prostate volume is used on men referred to urologist(2)

Table 21. Definitions of alternative biopsy methods

Biopsy method	Definition
Fusion	the patient undergoes a standard transrectal ultrasound guided biopsy but MRI targets from a preceding MRI scan are digitally "fused" to the ultrasound images so that additional cores can also be taken from those locations under ultrasound visualization [388]
Combined	both standard and targeted fusion biopsies are performed during a single biopsy session [151]
Cognitively guided	the patient undergoes a standard transrectal ultrasound guided biopsy, but the operator performs a biopsy on the basis of his or her knowledge of the location of the lesion at MR imaging [155]
In-gantry/In-bore	involves obtaining tissue samples with direct MR imaging guidance while the patient is in the MR imaging gantry and allows direct visualization of the MR imaging target and the needle at the same time [388]
Magnetic resonance spectroscopy imaging (MRSI)	Further to imaging of water and lipids, which is normally performed with MRI, MRS is a technique that provides detail on protons of molecules other than water and lipids. It can give quantitative information on the presence and quantity of metabolites in the prostate which can be used to estimate the presence and aggressiveness of cancer in

	prostate tissue [159]
Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI)	dynamically measures a bolus pass of an intravenously administrated MR contrast agent through the prostate. Has been shown to be of use in detection and staging of PC within a multiparametric protocol [159]
Diffusion-weighted magnetic resonance imaging (DW-MRI)	evaluates the microscopic mobility of water molecules in tissue. In addition to its value in the detection of cancer DW-MRI has also been shown to be a promising marker of tumour aggressiveness [159]
^{18}F-Choline PET/mpMRI	mpMRI with the addition of 18F-Choline PET. 18F-Choline is a radioactive substance being studied in positron emission tomography (PET) imaging to find certain types of cancer. 18F-choline gets taken up by cells in the body and more of it is taken up by cancer cells than by normal cells. A PET scanner is used to find which cells in the body have taken up 18F-choline. Also called 18F-fluoromethylcholine, 18F-FMCH, and fluorine F 18-fluoromethylcholine.[389]

Appendix 2: Search terms

Ovid - Medline

1. exp prostatic neoplasms/
2. (cancer adj3 (prostate or prostatic)).tw.
3. (carcinoma adj3 (prostate or prostatic)).tw.
4. (neoplas$ adj3 (prostate or prostatic)).tw.
5. (malignan$ adj3 (prostate or prostatic)).tw
6. (prostat$ adj3 (neoplasm$ or cancer or carcinoma or tumo?r$ or malignan$)).tw
7. 1 or 2 or 3 or 4 or 5 or 6
8. Prostate-Specific Antigen/
9 (prostate specific antigen or prostate-specific antigen or psa) tw
10. Mass screening/
11. (Screen$ or test$) tw
12. 8 or 9 or 10 or 11
13. exp “costs and cost analysis”/
14. (model adj3 (economic or cost)).tw.
15. (cost adj3 (effect$ or util$)).tw.
16. (economic adj3 (anal$ or eval$)).tw.
17. (natural history model) tw
18. (screen$ model$) tw
19. (disease progression model$) tw
20. 13 or14 or 15 or 16 or 17 or 18 or 19

21. 7 and 12 and 20
22. limit 12 to yr="2006-Current"

Ovid - EMBASE

1. prostatic neoplasms [not a MESH term]
2. exp prostate tumor/ [broader than cancer]
3. (cancer adj3 (prostate or prostatic)).tw.
4. (carcinoma adj3 (prostate or prostatic)).tw.
5. (neoplas$ adj3 (prostate or prostatic)).tw.
6. (malignan$ adj3 (prostate or prostatic)).tw
7. (prostat$ adj3 (neoplasm$ or cancer or carcinoma or tumo?r$ or malignan$)).tw
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. Prostate-Specific Antigen/
10 (prostate specific antigen or prostate-specific antigen or psa) tw
11. Mass screening/
12. (Screen* or test*) tw
13. 9 or 10 or 11 or 12
14. exp Economic evaluation/
15. (model adj3 (economic or cost)
16. (cost adj3 (effect$ or util$))
17. (economic adj3 (anal$ or eval$)
18. (natural history model) tw
19. (screen$ model$) tw
20. (disease progression model$) tw
21. 14 or 15 or 16 or 17 or 18 or 19 or 20
22. 8 and 13 and 21
23. limit 12 to yr="2006-Current"

Cochrane – NHS EED

#1 MeSH descriptor: [Prostatic Neoplasms]
#2 (prostat* NEAR/3 (neoplasm* or cancer or carcinoma or tumor* or tumour* or malignan*)):
#3 screen* or test*
#4 MeSH descriptor: [Mass screening]
#5 (prostate specific antigen or prostate-specific antigen or psa) tw
#6 Prostate-Specific Antigen/
#7 #1 or #2
#8 #3 or #4 or #5 or #6
#9 #7 and #8
#10 limit publication year from 2006 to 2016, in Economic Evaluations

Cochrane - HTA

#1 MeSH descriptor: [Prostatic Neoplasms]
#2 (prostat* NEAR/3 (neoplasm* or cancer or carcinoma or tumor* or tumour* or malignan*)):

#3 screen* or test*
#4 MeSH descriptor: [Mass screening]
#5 (prostate specific antigen or prostate-specific antigen or psa) tw
#6 Prostate-Specific Antigen/
#7 #1 or #2
#8 #3 or #4 or #5 or #6
#9 #7 and #8
#10 limit publication year from 2006 to 2016, in Technology Assessments

Appendix 3: Data extraction form

Data extraction form for systematic review of model-based economic evaluation methods in prostate cancer screening.

Title
Author
Year
Publication type
Study objective
Economic evaluation type
Population – age/ethnicity/prevalence of prostate cancer
Country
Strategies compared
Threshold for a positive result
Frequency of screening
Starting/stopping age
Types of biopsy
Types of treatment
Outcome measure
Model type
Justification for model type
Software used
Cycle Length
Justification for cycle length
Time horizon
Justification for time horizon
Sensitivity analysis (methods for incorporating uncertainty)
Evidence base for quality of life
Evidence base for resource use
Evidence base for adverse effects

Is overdiagnosis/overtreatment reported? If so, how?
Characterisation of disease (Stage or grade progression)
Evidence source for natural history pathway
Data on sensitivity/specificity
Data on clinical detection
Data on natural history
VOI (EVPPI or EVSI)
Cost-effectiveness result

Appendix 4: Reported accuracy of tests

Table 22. Reported accuracy of tests compared in studies

Study	Test	Sensitivity	Type of cancer sensitivity estimated for	Specificity	Type of cancer specificity estimated for	Source
Dijkstra et al 2017[147]	SelectMDx	0.957	high grade	0.336	low grade	Two prospective multicentre studies in the Netherlands in men who were scheduled for prostate biopsies, based on elevated PSA levels (≥3 ng/ml), abnormal DRE, or a family history of PCa (n=619)[164]
	SelectMDx			0.608	no cancer	
Bouttell et al 2019[140]	PHI cut off 25	0.887	any grade	0.365	any grade	Prospective cohort study of Chinese men with PSA 4-10 ng/mL and non-suspicious DRE (n=569)[390]
	PHI cut off 35	0.613	any grade	0.775	any grade	
	PHI cut off 55	0.129	any grade	0.974	any grade	
Sathianathen et al 2018[146]	TRUS guided biopsy	0.46	low grade			Taken from de Rooij CEA[156]
	TRUS guided biopsy	0.67	high grade			biopsy simulation study [391]

	PHI	0.883	low grade	0.294	no cancer	US multi-center, double-blind, case-control clinical trial in men with no history of PCa, non-suspicious DRE and pre-study PSA of 1.5–11.0 ng/mL (n=1372)[392]
	PHI	0.914	high grade			
	MRI	0.680	low grade	0.51	no cancer	
	MRI	0.880	high grade			
	4k score	0.816	low grade	0.3801	no cancer	US Multi-institutional Prospective Trial in men referred for biopsy (n=1012)[393]
	4k score	0.948	high grade			
	Select MDx	0.830	low grade	0.4	no cancer	Two prospective multicentre studies in the Netherlands in men who were scheduled for prostate biopsies, based on elevated PSA levels (≥3 ng/ml), abnormal DRE, or a family history of PCa (n=619)[164]
	Select MDx	0.910	high grade			
	EPI	0.800	low grade	0.3918	no cancer	Prospective multicenter study in the US in men with with PSA levels of

						2 to 20 ng/mL (n=255)[394]
	EPI	0.919	high grade			
	MRIGB	0.440	low grade			
	MRIGB	0.910	high grade			
Govers et al 2018[145] Govers et al 2019[139]	SelectMDx	0.957	high grade	0.61	No cancer	Two prospective multicentre studies in the Netherlands in men who were scheduled for prostate biopsies, based on elevated PSA levels (≥3 ng/ml), abnormal DRE, or a family history of PCa (n=619)[164]
		0.660	low grade	0.34	low grade	
Heijnsdijk et al 2016[148]	PSA test	0.790	T1 Gleason 2-6			ERSPC trial (n=42,376)[178]
		0.990	T3 Gleason 8-10			
	Biopsy	0.900				
Schiffer et al 2012[141]	UPA-PC	0.86		0.59		Prospective study carried out in Germany of men with suspicious PSA and/or DRE (n=211)[141]
	Biopsy	0.70		1		
Barnett et al 2018[151]	Standard Biopsy	0.8				Retrospective analysis of 7643 needle biopsies carried out in the US[395]

	Targeted fusion biopsy	0.770	high grade	0.68	high grade	Prospective cohort study of men with elevated PSA or abnormal DRE undergoing both targeted and standard biopsy concurrently at the National Cancer Institute in the US (n=1003)[180]
	Combined biopsy	0.850	high grade	0.49	high grade	
	PI-RADS > 3	0.965	clinically significant	0.597	clinically significant	Prospective study of men who presented for transperineal biopsy after mpMRI in one UK institution (n=201)[179]
	PI-RADS > 4	0.789	clinically significant	0.789	clinically significant	
Pahwa et al 2017[153]	MR Imaging	0.760		0.88		Taken from de Rooij CEA[156]
	Standard Biopsy	0.460				
	MR Cognitive biopsy	0.780	clinically significant			Two retrospective studies, one of 178 men and another of 54 men both undergoing MRI due to elevated PSA levels in Japan [396 397]
	MR Cognitive biopsy	0.200	clinically insignificant			Systematic review and meta-analysis of 16 studies including 1926 men[398]
	MR fusion biopsy	0.800	clinically significant			Retrospective analysis of men who underwent prebiopsy mpMRI followed by MRI fusion targeted biopsy in one

						US institution (n=452)[399]
	MR fusion biopsy	0.510	clinically insignificant			Systematic review and meta analysis of 16 studies including 1926 men[398]
	MR guided in-gantry biopsy	0.920	clinically significant			Single-institution, prospective study of biopsy-naive men referred to a urologist with elevated PSA in Australia (n=223)[400]
	MR guided in-gantry biopsy	0.210	clinically insignificant			Systematic review and meta analysis of 16 studies including 1926 men[398]
	All biopsy procedures			1		Assumption
Venderink et al 2017 [154]	mpMRI	0.930	clinically significant	0.21	clinically significant	Prospective cohort study of UK men with clinical suspicion of PCa who underwent mpMRI followed by TPM (n=129)[401]
	mpMRI			0.28	any cancer	
	MRI-TRUS fusion	0.770	clinically significant			Prospective cohort study of men with elevated PSA or abnormal DRE undergoing both targeted and standard
	MRI-TRUS fusion	0.500	clinically insignificant			

						biopsy concurrently at the National Cancer Institute in the US (n=1003)[180]
	MRI-TRUS fusion			1	any cancer	Assumption
	TRUS guided biopsy	0.530	clinically significant			Prospective cohort study of men with elevated PSA or abnormal DRE undergoing both targeted and standard biopsy concurrently at the National Cancer Institute in the US (n=1003)[180]
	TRUS guided biopsy	0.550	clinically insignificant			
	TRUS guided biopsy			1	any cancer	Assumption
de Rooij et al 2014[156]	TRUS guided biopsy	0.456		0.88		Sensitivity: Single institution retrospective study (n=438), Single institution prospective study (n=100), Single institution prospective study (n=54), Single institution prospective study (n=71) [402-404 405] Specificity: Single institution prospective study (n=64)[406]

	mpMRI	0.740		0.88		Meta-analysis of seven studies including 526 patients[407]
	MRGB	0.900		1		Assumption
Mowatt et al 2013[159]	TRUS guided biopsy	0.832		1		Prospective study of Italian patients suspected of harbouring PCa after a first negative biopsy (n=340)[408]
	T2-MRI	0.86		0.55		Systematic review and meta-analysis of 15 studies in 620 patients[159]
	MRS	0.92		0.76		Systematic review and meta-analysis of 10 studies in 438 patients[159]
	DCE-MRI	0.79		0.52		Systematic review and meta-analysis of 3 studies in 209 patients[159]
	T2-MRI or MRS	0.96		0.31		Systematic review and meta-analysis of 8 studies in 316 patients[159]
	T2-MRI or DCE-MRI	0.88		0.14		Systematic review and meta-analysis of 3 studies in 209 patients[159]

Barnett et al [150]	mpMRI alone (Likert 4–5)	0.8519 (0.6627–0.9581)		0.5517 (0.3569–0.7355)		US prospective single-arm clinical trial (N=63)[409]
	mpMRI alone (PI-RADSv2 3–5)	0.9259 (0.7571–0.9909)		0.5862 (0.3894–0.7648)		
	18F-choline PET/mpMRI (Likert)	0.9259 (0.7571–0.9909)		0.9310 (0.7723–0.9915)		
	18F-choline PET/mpMRI (PI-RADSv2)	0.8889 (0.7084–0.9765)		0.9310 (0.7723–0.9915)		
	Standard biopsy	0.80				Prospective study of autopsy prostates from 164 men who had no history of prostate cancer[410]
	Combined biopsy (targeted biopsy and standard 12-	0.85	Gleason score of ≥ 3 + 4	0.49	Gleason score of ≥ 3 + 4	Prospective cohort study of 1003 men undergoing both targeted and standard biopsy concurrently from 2007 through 2014 at the National Cancer Institute in the US [180]
Kim et al [142]	PHI > 20	0.99	≥G2	0.10	≥G2	UK prospective five-centre study (N=545)
	PHI > 20	1.00	≥CPG3	0.08	≥CPG3	
	PHI > 25	0.96	≥G2	0.25	≥G2	
	PHI > 25	0.99	≥CPG3	0.22	≥CPG3	
	PHI > 30	0.92	≥G2	0.40	≥G2	
	PHI > 30	0.95	≥CPG3	0.35	≥CPG3	

	PHI > 35	0.87	≥G2	0.55	≥G2	
	PHI > 35	0.93	≥CPG3	0.49	≥CPG3	
	PSA density > 0.10	0.93	≥G2	0.31	≥G2	
	PSA density > 0.10	0.97	≥CPG3	0.28	≥CPG3	
	PSA density > 0.15	0.81	≥G2	0.51	≥G2	
	PSA density > 0.15	0.90	≥CPG3	0.53	≥CPG3	
	PSA density > 0.20	0.69	≥G2	0.77	≥G2	
	PSA density > 0.20	0.80	≥CPG3	0.72	≥CPG3	

Legend: PSA: Prostate Specific Antigen, DRE: Digital Rectal Examination, PCa: Prostate cancer, PHI: Prostate Health Index, TRUS : Transrectal Ultrasound Guided, CEA: Cost-effectiveness Analysis, MRI: Magnetic Resonance Imaging, MRIGB: MRI Guided Biopsy, mpMRI: Multiparametric MRI, EPI: ExoDx Prostate (IntelliScore), ERSPC: European Randomized study of Screening for Prostate Cancer, UPA-PC: Urinary Proteome Analysis for PCa diagnosis, TPM: Template Mapping Biopsy, DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging, DW-MRI: Diffusion-weighted magnetic resonance imaging, CPG3: Cambridge Prognostic Group 3

Appendix 5: CHEERS and ECOBIAS forms

Table 1. CHEERS criteria met

Item	Item No	Recommendation	Study reference number																					
			139	264	411	142	379	140	157	155	412	147	149	148	413	299	145	152	153	156	151	146	150	28
Title and abstract																								
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	1	1	1	1	1	1	0	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Introduction																								
Background and objectives	3	Provide an explicit statement of the broader context for the study.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
		Present the study question and its relevance for health policy or practice decisions.	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Methods																								
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	0	1	0	0	0	1	1	1	0	1	0	0	1	1	1	1	0	1	0	1	1	1
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	1	1	1	0	1	0	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	1	1	1	0	1	2	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	1	1	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	0	0	0	1	2	0	2	0	1	0	0	2	2	0	0	1	2	2	0	2	1	2
	11b	*Synthesis-based estimates*: Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	2	0	2	2	1	2	1	2	2	2	2	1	1	2	2	1	0	0	2	0	2	1
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	1	0	0	0	1	1	1	0	0	1	1	0	1	1	1	1	1	0	0	0	1	1
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	1	1	0	0	0	2	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	1	1	1	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0	1	1
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data;	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1

		extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.																							
Results																									
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	0	1	1	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	0	1
Characterising uncertainty	20a	*Single study-based economic evaluation:*Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	0	1	1	0	1	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	

Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	2	2	2
Discussion																									
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Other																									
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	1	1	0	1	1	1	1	0	1	0	1	0	1	0	1	1	1	1	1	0	1	0
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	0	1	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1
Mean percentage of applicable criteria met			65	80	68	50	84	59	89	63	37	79	63	58	100	74	68	90	89	74	53	68	74	844

Table 2. ECOBIAS checklist

	Govers et al[139]	Karlsson et al[264]	Teoh et al [411]	Kim et al [142]	Nicholson et al [379]	Boutell et al[140]	NICE guideline[157]	Cerantola et al [155]	Schiffer et al[412]	Dijkstra et al[147]	Nichol et al[149]	Heijnsdijk et al[148]	Mowatt et al[413]	Venderink et al[299]	Govers et al[145]	Faria et al[152]	Pahwa et al[153]	de Rooij et al[156]	Barnett et al[151]	Sathianathen et al[146]	Barnett et al [150]	Callner et al[28]
Narrow perspective bias	No	Yes	No	No	No	No	No	No	No	No	Yes	Yes	No	No	No	No	Unclear	No	No	No	No	No
Inefficient comparator bias*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cost measurement omission bias	Yes	Yes	Yes	Yes	Yes	Partly	Yes	Yes	No	Partly	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Partly	Yes	Partly	Yes	Yes
Intermittent data collection bias	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	Yes	NA	NA	NA	NA	NA	NA	NA	NA
Invalid valuation bias	Partly	Partly	No	Partly	Yes	No	Yes	Partly	No	No	Partly	No	Yes	Yes	No	No	Yes	No	No	Partly	Yes	Yes
Ordinal ICER bias	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double-counting bias	No	No	No	Yes	No	No	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Inappropriate discounting bias	Yes	Yes	Yes	No	Yes	NA	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Limited sensitivity analysis bias§	No	Partly	No	No	Partly	No	Partly	No	No	No	No	No	Yes	No	No	No	Partly	No	No	No	No	No
Sponsor bias	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Reporting and dissemination bias	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	No	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Structural assumptions bias	No	Yes	Yes	Partly	Yes	Yes	Yes	Partly	Partly	Partly	No	Yes	Yes	Partly	Partly	Partly	Partly	Partly	Yes	Partly	Yes	Partly
No treatment comparator bias*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wrong model bias	Partly	Yes	Partly	Yes	Yes	Partly	Yes	Partly	Partly	Yes	Partly	Yes	Yes	Yes	Partly	Yes	Yes	Yes	Partly	Partly	Yes	Partly
Limited time horizon bias	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Unc	Yes

																					lear	
Bias related to data identification	Partly	Partly	No	Partly	Partly	No	Yes	Partly	Partly	Partly	Partly	Partly	Yes	Partly	No	Yes	Yes	Partly	Partly	Partly	Partly	Partly
Bias related to baseline data	NA	Yes	Unclear	Yes	Yes	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Yes
Bias related to treatment effects	NA	NA	NA	NA	NA	NA	Yes	NA	NA	NA	NA	NA	Yes	NA	NA	NA	NA	NA	NA	NA	NA	NA
Bias related to quality-of-life weights (utilities)	Yes	Yes	Yes	NA	Yes	NA	Yes	Yes	NA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Non-transparent data incorporation bias	Yes	Yes	No	Partly	Yes	No	Yes	Partly	Yes	Partly	Yes	No	Yes	Partly	No	Partly	Yes	Yes	Partly	Partly	Partly	Yes
Limited scope bias§	No	Partly	No	No	Partly	No	Partly	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No
Bias related to internal consistency	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear

Appendix 6: Participant Information Sheet

Gaining consensus on UK relevant prostate cancer screening strategies

Information Sheet for Participants

A team of researchers at the University of Bristol are inviting you to take part in a research study. This study will help us understand the prostate cancer screening strategies that are relevant to compare in an economic model, given the rapidly changing landscape in prostate cancer screening. The study will involve you completing two rounds of questionnaires, roughly one month apart, in which you indicate your preferred screening strategy.

What is the aim of this study?

The first step in any economic evaluation is to determine the decision question, which includes identifying all relevant comparators or in this case screening strategies. As prostate cancer screening is a rapidly developing area of research, the aim of this piece of work is to elicit the most relevant strategies to compare in an economic model from the experts working in the field.

This is important as there is currently no screening programme for prostate cancer in the UK, due to the benefits having not been shown to outweigh the potential harms. Current findings indicate that even if population screening for

prostate cancer, starting with a simple PSA test, can prevent death from the cancer for a subset of men, this is accompanied by a significant amount of overdiagnosis and subsequent overtreatment. These unnecessary tests and treatments are costly both economically on healthcare resources and in harm caused to men.

Recent developments such as the potential for risk stratification using genetic risk scores, potentially better biomarkers e.g. STHLM3, and better diagnostic strategies (including the use of multiparametric MRI), offer opportunities for improving the outcomes in screening, particularly reduction in overdiagnosis and higher specificity for potentially lethal cancer. However, it is not yet clear how or if they should be implemented. In addition, there is uncertainty as to the age at which to start and stop screening, how often men should be screened, and the factors that should indicate further investigation via biopsy. An evidence dossier summarising these issues accompanies this information sheet.

Why have I been asked to take part?

You have been asked to take part in the study because you are an expert in prostate cancer screening. We would like to know which screening strategies you think it would be useful to have economic evidence on in terms of lifetime costs and effects. The findings from the research will help decision-makers to decide if or how prostate cancer screening should be provided in the UK.

What will I have to do?

Round 1

The first step will be for you to complete the attached questionnaire which asks questions relating to different aspects of prostate cancer screening e.g. who should be screened and how. It is recommended you read the evidence dossier before completing the questionnaire. When filling out the

questionnaire, we strongly encourage you to provide as much detail as possible in the comments section after each question on why you chose the option you did or any issues or concerns you may have with the question. If your preferred option is not included, please suggest it using the 'other' option. At the end of the questionnaire you will see a personalised statement summarising your responses.

At this stage, we would greatly appreciate if you could also suggest others (in the form of names and email addresses) within or outside your organisation who may also be willing to complete the questionnaire and who may offer additional insight.

Round 2

You will be sent a modified questionnaire where items not chosen by any participant in round 1 are removed and items suggested by participants in round 1 are added. Feedback from round 1 will be presented for each round 2 item in the form of the percentage of participants choosing each option in round 1 along with a reminder of your own choice. Comments from round 1 will also be summarized and presented.

You will be asked to recomplete the questionnaire taking into account the information provided.

Do I have to take part?

No, you don't have to take part. If you decide to take part and then change your mind, you can withdraw from the research at any time. If you decide to withdraw from the research, please ensure to notify us.

What will happen to the study results?

The screening strategies identified in this process will be compared to each other and to opportunistic screening in a lifetime economic model. The results

of this consensus process and of the economic modelling will be published in peer-reviewed journals and presented at conferences.

What will happen to your data?

Your data will be analysed and stored electronically on a secure computer network at the University of Bristol. Your personal data will be stored separately from research findings and will only be accessed by the research team. Any information that could identify you will be removed from the data before the findings are seen by others, and personal data will not be used in research reports. The handling, processing, storage, and destruction of these data are compliant with the Data Protection Act 1998. At the end of the study your research data will be made "Controlled Access", which means your data will be stored in an online database and can be accessed and used by other researchers through requests to a Data Access Committee. However, there will be no way to identify you from these data.

What do I do next?

If you would like to take part in the study, please complete and return the questionnaire which is linked to in this email. Please also suggest any other experts in prostate cancer screening who may be willing to take part.

How this study is funded

The study is funded by Cancer Research UK and has been approved by the University of Bristol Ethics Committee.

Research team

Edna Keeney, PhD student and Senior Research Associate in Statistical and Health Economic Modelling
Sabina Sanghera, Lecturer in Health Economics
Howard Thom, Lecturer in Health Economics
Emma Turner, Research Fellow
Richard Martin, Professor of Clinical Epidemiology

All affiliated to:
Population Health Sciences, Bristol Medical School, University of Bristol

Can I have more information?

If you have any questions, please feel free to contact the research team using the details below:

Edna Keeney

Population Health Sciences, Bristol Medical School
University of Bristol, 1-5 Whiteladies Road
Bristol, BS8 1NU
Telephone: 0117 42 83118
Email: edna.keeney@bristol.ac.uk

Appendix 7: Example of completed questionnaire

Confidential

Record ID 33
Page 1

Gaining consensus on prostate cancer screening strategies

Please complete the survey below.

Thank you!

Response was added on 11-02-2020 16:46.

What is your email address?	x@ymail.com
What is your age range?	◯ 18-24 ◯ 25-34 ◯ 35-44 ⊗ 45-54 ◯ 55-64 ◯ 65-74 ◯ 75+
How many years have you been working in your field?	◯ 1-5 ◯ 5-10 ⊗ 10-15 ◯ 15-20 ◯ 20+
What is your clinical focus?	◯ Oncology ⊗ Urology ◯ General practice ◯ Other
In what country are you based?	UK

1a. What type of prostate cancer screening programme do you feel should be provided in the UK? Note: Age-based implies that an invitation to screening will depend on age alone. Risk-based implies that an invitation will depend on both age and level of risk.	◯ No screening ◯ Opportunistic screening ⊗ Organised screening - Age-based ◯ Organised screening - Risk-based ◯ Other
1b. Comments on this question/Reason for your answer	_____
2a. If age-based screening were to be provided at what age should a man's baseline PSA level should be taken?	◯ 40 ◯ 45 ⊗ 50 ◯ 55 ◯ 60 ◯ Other
2b. Comments on this question/Reason for your answer	_____
3a. If age-based screening were to be provided at what age should screening end?	◯ 70 ⊗ 75 ◯ 80 ◯ Other
3b. Comments on this question/Reason for your answer	_____
9a. Which PSA threshold do you think should be used to indicate further investigation?	⊗ 3 ng/ml ◯ 4 ng/ml ◯ 5 ng/ml ◯ 6 ng/ml ◯ 7 ng/ml ◯ 8.5 ng/ml ◯ 10 ng/ml ◯ This should be based on age ◯ Other
9b. Comments on this question/Reason for your answer	_____
10a. Assuming some optimal strategy for inviting men to be screened has been adopted, how frequently do you think men should be screened?	◯ every 10 years ◯ every 6 years ◯ every 4 years ⊗ every 2 years ◯ annually ◯ only once ◯ This should be based on PSA level ◯ Other

10b. Comments on this question/Reason for your answer	_____
11a. What further investigation should men with a raised PSA level have prior to being offered a biopsy?	☐ no further investigation ☐ Digital Rectal Examination (DRE) ☒ multi-kallikrein panel (e.g. 4k score, STHLM3) ☐ PSA density ☐ % free PSA ☒ multiparametric MRI (mpMRI) ☐ other
11b. Comments on this question/Reason for your answer	_____
12a. What is the main factor you would consider in your decision on whether or not a screening programme should be introduced?	◯ false findings ◯ reducing death from prostate cancer ◯ quality of life ◯ complications ◯ workforce capacity ◯ financial resources ◯ cost of testing ⊗ cost-effectiveness of screening ◯ other
12b. Comments on this question/Reason for your answer	_____

Your preferred screening strategy is

Organised screening - Age-based

In men aged 50 to 75.

The PSA threshold for further investigation should be 3 ng/ml.

Men should be screened every 2 years.

Men with a raised PSA should be tested using multi-kallikrein panel (e.g. 4k score, STHLM3), multiparametric MRI (mpMRI) prior to biopsy.

The main factor you would consider in your decision to screen is cost-effectiveness of screening.