Evaluation of the Cost-Effectiveness of Novel Tests in the Screening and Diagnostic Phases of Prostate Cancer Compared to Current Practice

Master's Thesis in
Biomedical Engineering & Mathematical Statistics

Ebba Palenius & Elin Bonnevier

Supervisor, McGill University:
Dr. Alice Dragomir

Supervisors, Lund University:
Andreas Jakobsson
Johan Malm

Faculty of Engineering

Carried out at Department of Biomedical Engineering &
Centre for Mathematical Sciences

August 2017

Abstract

Introduction There are major limitations in the current methods for screening and diagnosis of prostate cancer. In this project, the implementation of new tests has been researched. The assessed tests in the screening phase are the 4Kscore®, Prostate Health Index (PHI), Magnetic Resonance Imaging (MRI) and Prostarix. The assessed tests for use after a negative biopsy are Prostate Cancer gen 3 (PCA3), ConfirmMDx, Prostate Core Mitomic Test (PCMT), PHI and the 4Kscore®. The aim was to evaluate the cost-effectiveness of introducing these tests compared to the current use of Transrectal Ultrasound-Guided Biopsies (TRUSGB) or MRI-Guided Biopsies (MRGB).

Methodology A systematic literature review was conducted to collect all available clinical utility studies regarding the tests. Markov models were developed based on a previously published TRUSGB model to estimate Quality-Adjusted Life Years (QALY) gained and costs for the tests over 5, 10, 15 and 20 years.

Results The resulting models were for 4Kscore®, PHI, PCA3 and ConfirmMDx. Compared to the current method 4Kscore® and PHI showed better results while ConfirmMDx and PCA3 were equal. Only 4Kscore® showed better potential than the MRGB method.

Conclusion Compared to the TRUSGB and MRGB methods, implementation of the
4Kscore® test was the dominating strategy. More clinical utility studies are needed to confirm the results.

Keywords: Prostate Cancer; Cost-Effectiveness; Screening; Clinical Utility; Markov modeling; Biopsy; TreeAge

Preface

This Master's thesis finalizes five years of studies in Biomedical Engineering and Risk Management at the Faculty of Engineering, Lund University. The thesis was conducted at the Division of Urology at the Research Institute of the McGill University Health Centre (MUHC) in Montréal, Canada. The work was done from February 2017 until July 2017.

Numerous people have supported us during our studies and during our time in Montréal, we can not thank all but we are very grateful. We would like to express our appreciations especially to the following persons, without given order, who have been a support during the past months:

Dr. Alice Dragomir, our supervisor at the Research Institute, who provided us with the project and supported us in our work with guidance and her knowledge. Thank you for inviting us to your team and for making us feel at home. We would also like to thank our supervisors at Lund University, Johan Malm and Andreas Jakobsson, who helped us from Sweden when questions arose along the way.

We would like to thank Ghadeer Olleik and Abdel Tarifi, it was a pleasure to perform the systematic literature review in collaboration with both of you. Many thanks to the other participants in the team: Halima Lahcene, Sara Nazha, Jason Hu, Noemie Provost and Emma Nablsi, who made our time in Montréal memorable and who were always happy to encourage or help us if we needed anything.

We would also like to thank Thomas Laurell and Anna Lindgren for being examiners of this Master's thesis.

Last, but not least, thanks to family and friends for always believing in us and being there for us.

Acronyms

ADT Androgen Deprivation Therapy.

AS Active Surveillance.

cPSA complexed Prostate Specific Antigen.

CRPC Castration-Resistant Prostate Cancer.

DRE Digital Rectal Exam.

EBRT External Beam Radiation Therapy.

ERSPC European Randomized Study of Screening for Prostate Cancer.

FDA U. S. Food and Drug Administration.

fPSA free Prostate Specific Antigen.

FU Follow Up.

hK2 human Kallikrein-2.

IMRT Intensity-Modulated Radiation Therapy.

MRGB Magnetic Resonance imaging-Guided Biopsy.

MRI Magnetic Resonance Imaging.

MUHC McGill University Health Centre.

PCA3 Prostate Cancer gen 3.

PCMT Prostate Core Mitomic Test.

PHI Prostate Health Index.

PLCO Prostate, Lung, Colorectal, and Ovarian Cancer screening trial.

PSA Prostate Specific Antigen.

QALY Quality-Adjusted Life Years.

RP Radical Prostatectomy.

RT-PCR Reverse Transcription Polymerase Chain Reaction.

TRUSGB Transrectal Ultrasound-Guided Biopsy.

USPSTF United States Preventive Services Task Force.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Background

Prostate cancer is one of the most commonly diagnosed cancers in men today, with approximately one in seven American men being diagnosed in their lifetime [1]. Although it is a very common disease, the mortality is low, with a death rate of 16% among diagnosed men [2, p. 2771]. Because of the aging population today and the fact that the average age of diagnosis is 67 years [2, p. 2763], the occurrence of prostate cancer will most likely continue to rise [2, p. 2771]. In the future this can lead to the disease becoming a bigger health concern in society than it is today.

Prostate cancer usually progresses slowly and due to the high average age of the patients, many die from other causes before symptoms appear. Early stages of prostate cancer rarely cause any symptoms and usually the symptoms do not appear until the cancer has advanced and spread to other parts of the body (metastasized) [2, p. 2763]. A high proportion of the patients would never have discovered the disease if it had not been detected during screening [2, p. 2763]. In these cases, the benefits with treatment might not outweigh the cost and potential complications. This is the case especially among older men for whom treatment is associated with minimal benefit and a higher risk. Due to the high average age at diagnosis, many patients would probably have had a higher quality of life without treatment, and can thus be considered as overdiagnosed and overtreated if treated [2, p. 2763].

If the cancers are found early, there is a greater chance to cure the disease. The patients may also avoid aggressive treatments and potential suffering. Some tumors are aggressive, spread outside the prostate and cause discomfort and eventually death. These are the cancer cases that need to be identified and treated as early as possible.

Prostate Specific Antigen (PSA) is an antigen that has been used in screening for prostate cancer since the end of the twentieth century [2, p. 2560]. The screening method has led to increased early-stage cancer detection but

as mentioned before, this can lead to overdiagnosis and overtreatment. Another problem with PSA screening is that the marker is prostate specific, not prostate cancer specific [3]. An increased PSA value can be due to other, benign, prostate conditions that are hard to distinguish from cancer.

In 2012, United States Preventive Services Task Force (USPSTF) stated that the benefits of PSA-based screening for prostate cancer do not outweigh the harms of prostate biopsy for any men. This statement was based on two randomized controlled trials published in 2009 that questioned the utility of PSA in screening. The Prostate, Lung, Colorectal, and Ovarian Cancer screening trial (PLCO) found no difference in the rate of death from prostate cancer between men included and excluded in PSA screening [4]. In the European Randomized Study of Screening for Prostate Cancer (ERSPC) the result was a significant decline in death from prostate cancer in the screening group, but over 1400 men needed to be screened in order to prevent one death [4]. Due to this controversy, more studies are needed to prove the utility of screening for prostate cancer. Also, new tests that have a potential to improve the screening process need to be researched.

The low specificity of existing tests and the fear of missing significant prostate cancer results in a high proportion of biopsies that are considered as unnecessary. Physicians are often presented with the dilemma that patients undergo several negative biopsies but still have an increased PSA value or other suspicions of prostate cancer. Prostate biopsies may cause pain, discomfort, anxiety or other complications that are related to costs to the health care system. Therefore, it is important to minimize the number of unnecessary biopsies without compromising on the specificity.

An increasing number of options are available in the health care and due to the nature of the consequences, decisions are difficult to make. There is not only a question of curing the disease, but also to improve the condition or increase the quality of life. This progress has helped a lot of people but it also makes the decision making complex for clinicians. The human capacity to handle complex problems is limited and it is difficult to focus on many aspects at once. New more reliable tests could help the clinicians in deciding what patients are in risk of prostate cancer and to stratify them into risk groups. Modeling the disease can give results depending on many variables and illustrate the theoretical utility of introducing new tests [5]. There is a need for more research and new tests that can improve the screening process to reduce over-diagnosis and unnecessary biopsies. The research on new screening methods for prostate cancer is limited and proof of cost-effectiveness is needed.

One study has been previously performed and published by a team at the Research Institute of the McGill University Health Centre (MUHC) to inves-

tigate the cost-effectiveness of Magnetic Resonance imaging-Guided Biopsy (MRGB) compared to the conventional method with Transrectal Ultrasound-Guided Biopsy (TRUSGB) [6]. This study has been the base of the work in this project and is hereafter referred to as the previous study.

This project was a part of a bigger study that was performed by a local team at the Research Institute of the MUHC under supervision of Dr. Alice Dragomir. Some work was performed in collaboration with the local team, consisting of Ghadeer Olleik, Abdel Tarifi, Halima Lahcene, Sara Nazha, Jason Hu and Noemie Provost. This is described in detailed in the report.

1.2 Objectives

The specific objectives of this project were to:

- Conduct a systematic literature review to collect all available clinical utility studies of high quality regarding new tests in the screening phase or the phase after a negative biopsy in prostate cancer. Evaluate interventions or management strategies with highest evidence of clinical outcomes that are most likely to be adopted in clinical practice.
- Develop predictive Markov models covering evolution and management of prostate cancer for the identified strategies from screening and diagnosis until end-of-life.
- Estimate the cost-effectiveness of the identified strategies and evaluate them in comparison to the results from the Markov modelling from the previous study.

1.3 Limitations

- Only articles written in English or French were evaluated.
- Conference abstracts were not assessed.
- Relevant articles might have been missed. A systematic literature review includes all available studies on a subject. This was the goal in this project but it is impossible to know if some important articles were missed due to the limited number of databases that were researched or the search terms that were chosen.
- The total time of the project limited the time spent on the systematic literature review and modeling.

1.4 Division of work

All work was discussed and reviewed continuously by the authors. However, E. Bonnevier was responsible for researching the screening tests while E. Palenius was responsible for the tests performed after a negative biopsy. This was the case both during the systematic review and modeling. During the final phase of the project, E. Palenius focused mostly on the medical background in the report while E. Bonnevier was responsible for analyzing the results from the model. Both authors contributed equally to all parts of the project.

The work done in collaboration with members of the local team is presented and described in detail in the methodology section.

1.5 Structure of this report

This report is divided into 6 chapters. The first chapter gives a introduction to the issues regarding screening for prostate cancer and decision making in clinical practice. This chapter also includes the objectives of this project presented along with limitations and the division of work.

Chapter two includes the theory behind the project. It includes theory about prostate cancer, current practice in the health care, the new evaluated methods along with theory about the systematic literature review and health economical modeling.

The third chapter describes how the project was performed. The methodology of the systematic literature review and the modeling is presented in detail.

The results are presented in chapter four and discussed in chapter five along with discussions about the methodology, sustainability and ethical aspects as well as future work. The final chapter presents the conclusions from the project.

2 Theory

2.1 The prostate and prostate cancer

The prostate is a gland in men producing fluid that carries semen during ejaculation. It is normally the size of a walnut and is located under the bladder, in front of the rectum, and surrounds the urethra [7].

Normal cell growth in the prostate is controlled by male hormones. The prostate grows rapidly during puberty and keeps growing continuously until middle age. In the early middle age, the risk of rapid enlargement of the prostate is increased [8]. Disturbed balance between cell growth and cell death, leading to uncontrolled cell growth, can be a sign of cancer. Groups of cells might then form tumours which will either grow locally or metastasize to other organs. Prostate cancer often grows without symptoms and 80% of all cases are detected during routine medical checkups. If the disease is diagnosed in its early stages, it is often curable while symptoms can be eased and life be prolonged if the disease is more advanced [7]. Cancer that is not found in its early stages may become a locally advanced or metastatic prostate cancer. Common symptoms of these stages include bone pain, pathological fractures, anemia and leg swelling [2, p. 2763]. An illustration of a cancerous prostate and its location in the body can be seen in Figure 2.1.

Prostate cancer is one of the most common malignancies worldwide. The diagnosis is more common among older men and only 2% of the diagnoses are in men younger than 50 years old [2, p. 2706]. The knowledge in risk factors is limited but the disease is associated with high age and a family history of prostate cancer [7]. There are also geographical and ethnical variations in the prostate cancer incidence [2, p. 2708]. The lowest yearly incidence rates are seen in Asia and the highest in Scandinavia and North America, especially among African-American men. The rates differ more than 100-fold between countries with low and high incidence rates. These differences are not only in the incidence of the disease, but also in disease-specific mortality [2, p. 2706]. Although the differences are confirmed in several studies, the reason is unknown.

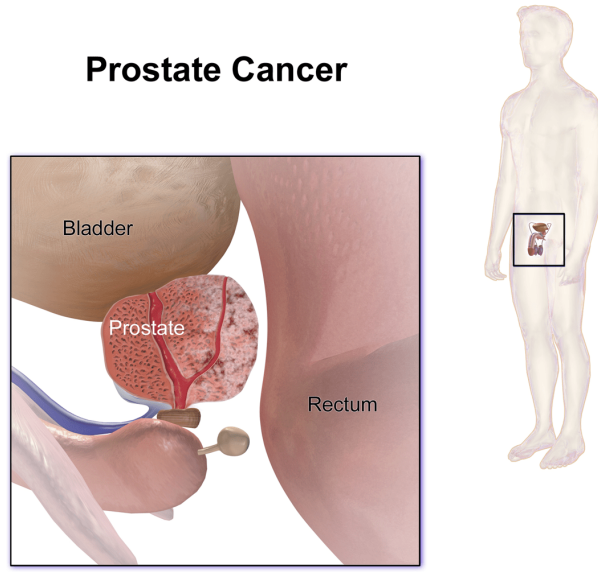


Figure 2.1: Location of the prostate, under the bladder in front of the rectum, in a patient with prostate cancer [9].

The commonly used categories such as African-American, white and Asian have social and cultural differences rather than biological ones. The differences may be due to variations in environment, diet, lifestyle and attitudes toward health care rather than genetic structure. Access to and quality of health care, accuracy of the cancer registries and the use of PSA screening also varies highly and affects the statistics of the disease [2, p. 2705]. The differences might also be connected to knowledge about the disease. For example, the reason for the higher incidence might be that men with a diagnosed relative are more likely to seek prostate cancer screening.

2.2 Current practice

There are three main clinically available methods today to diagnose prostate cancer: the Digital Rectal Exam (DRE), PSA blood test and prostate biopsies [7, p. 30]. DRE and prostate biopsies have been in use for a long time while PSA was introduced in the 1990s, with a controversial effect on incidence and mortality rates [2, p. 2735]. DRE and PSA are simple examinations that indicate a risk for prostate cancer, but a biopsy needs to be performed to determine the diagnosis. To grade the aggressiveness of an eventual cancer, the Gleason grading system is the most common method in use today.

2.2.1 Digital Rectal Examination

Before using PSA, DRE was the only method for early detection of prostate cancer, but today they are often used in combination [2, p. 2764]. The exam consists of a doctor palpating the prostate to detect abnormalities by inserting a gloved finger into the patients rectum.

The exam is limited since it is not possible to reach the whole prostate and a substantial proportion of early cancers are missed [2, p. 2764]. The exam has a lack of reproducibility and thus it often overestimates or underestimates the extent of the disease [2, p. 2768]. But since tumours often develop close to the rectum, many abnormalities can be detected and it is an important basic exam [7].

2.2.2 Prostate Biopsy

Today, the PSA test result and a DRE are used in the decision making process to determine if patients are in need of a biopsy or not.

The location of the prostate, in front of the rectum, is ideal for transrectal biopsies and imaging. The most commonly used biopsy is the TRUSGB where an ultrasound probe is inserted into the rectum to create images of the prostate tissue and guidance in the decision of the exact placement of the biopsy [2, p. 2737], see Figure 2.2. During a TRUSGB, a number of samples are taken from the prostate. The cores are analyzed in microscope by a physician to determine presence of cancerous tissue and determine the aggressiveness of the cancer. If cancerous cells are found it is counted as a positive biopsy and otherwise a negative biopsy. Another biopsy method in use today is MRGB, where the placement of the tumour is determined with another common imaging technique, Magnetic Resonance Imaging (MRI).

TRUSGB is the most common way of diagnosing prostate cancer today but the method has several limitations [2, p. 2737]. A high proportion of the results are false-negative, mainly because only small parts of the prostate are tested in every biopsy. If the initial biopsy is negative or indeterminate, there is often still a suspicion of prostate cancer. In this case additional biopsies can be performed, but there is no guarantee that these will catch the eventual cancer either. A positive result on the biopsy can also be misleading because of the chance that it is just a small, low risk cancer that does not affect the patient and does not need any treatment. The diagnosis of this cancer may cause anxiety for the patient, in addition to the complications after a biopsy such as pain and bleeding.

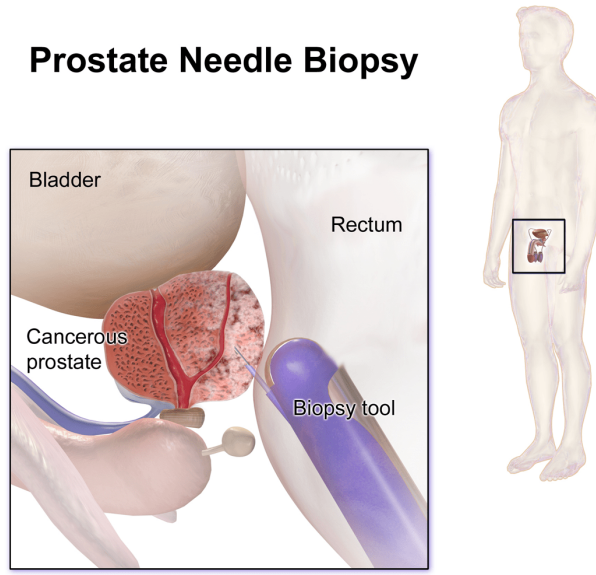


Figure 2.2: Transrectal Ultrasound-Guided Biopsy of a cancerous prostate [10].

2.2.3 The Gleason Grading system

There are several available methods to grade the aggressiveness of prostate cancer tumours after a positive biopsy, but the most widely accepted is the Gleason grading system. It was first tested in the middle of the twentieth century by an American pathologist named Dr. Donald F. Gleason together with members of the Veterans Administration Cooperative Urological Research Group (VACURG) and it is the most frequently used method for grading prostate cancer tumours worldwide [11]. The Gleason grading system is based on the histological pattern of the cancer cells in biopsy tissue samples. The growth patterns were divided into five grades by Dr. Gleason, as illustrated in Figure 2.3.

The Gleason score is obtained by adding the grades of the two most common growth patterns in the tissue sample from the performed biopsy. If a tumour only consists of one growth pattern, the primary and secondary patterns are given the same grade. The score can range from 2 to 10 where a higher score corresponds to more differentiated cells which in many cases implicates a more aggressive cancer. In this project, low risk cancers are defined as Gleason score 2–6 while cancers of grade 7–10 are of intermediate/high risk. This stratification is used since it was defined in the previously performed study and since it was the most commonly used in the articles in the systematic literature review.

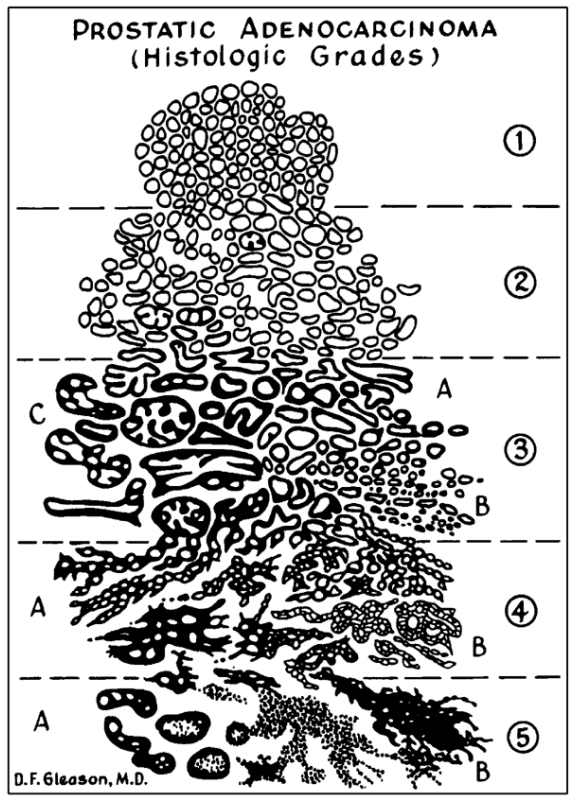


Figure 2.3: Cancer cell growth patterns sorted into the five Gleason grades [11].

2.2.4 Prostate Specific Antigen

PSA is measured in a blood test that can aid clinicians in their decision to perform a biopsy. It is a glycoprotein in the human kallikrein family that is produced by the prostate cells [8]. It is mainly found in the epithelial cells of the prostate and has been used as a marker for prostate cancer since 1988 [2, p. 2560].

PSA is normally present in low concentration in sera, both in unbound form and bound to proteins such as the antiproteases ACT and macroglobulin [2, p. 2751]. High levels of PSA can be detected in the blood if the prostate tissue is cancerous but not all men with prostate cancer have elevated PSA levels. The elevations can also be due to other prostate conditions, such as noncancerous (benign) prostate enlargement [7]. PSA levels also vary with age, race and prostate volume. It has been shown that not only the present risk but also the future risk of prostate cancer and the chance of finding

the disease on a biopsy increases with the PSA level [2, p. 2764]. In addition, the PSA level can increase after manipulations such as biopsies and DRE [2, p. 2752]. Since the PSA test is organ specific, not cancer specific, there is an overlap in concentration between benign and malignant disease. This makes the test very unspecific, but it can indicate a probability of having prostate cancer.

Even though PSA levels often rise in men with prostate cancer or other prostate diseases, prostate cancer cells normally produce less rather than more PSA than other cells. The elevated concentrations are probably due to cancer progression and destabilization of the prostate histological structure. The loss of barriers within the gland probably leads to PSA leaking into the circulation [2, p. 2751].

The PSA value can guide the clinicians in stratifying the patients in risk groups, but studies have shown that there is no PSA threshold that can rule out prostate cancer in any age range [2, p. 2740]. Healthy patients aged 50 to 80 years normally have PSA values that range from 1,0 to 4,0 ng/l [2, p. 2561]. Patients with a PSA value over 4 ng/ml are recommended a prostate biopsy to outrule a diagnosis by most clinicians, but a lot of research is done to investigate the optimal cut-off further. Patients with PSA values between 4 and 10 ng/ml are often the ones considered as being in the grey zone and would need to be investigated further with new methods [12], but the result is not a guarantee.

PSA derivatives

To increase the performance of the PSA measurements, derivatives have been researched. These include PSA density, PSA velocity and percentage of the ratio of free to total PSA (%fPSA) [2, p. 2740].

PSA density is calculated by dividing the PSA level with the total prostate volume. This measurement has shown utility in stratifying between patients with prostate cancer and benign prostatic hyperplasia and has been associated with cancer aggressiveness [2, p. 2766]. PSA velocity is the evaluation of rate of change in PSA. Fluctuations in PSA can occur due to various reasons but a rising PSA level corrected for the time between measurements has been connected to prostate cancer [2, p. 2765].

As mentioned earlier, PSA is found in the circulation in small concentrations, both complexed and free. New antibodies that are specific to free Prostate Specific Antigen (fPSA) and complexed Prostate Specific Antigen (cPSA) have made it possible to measure the different forms and their ratios. Patients with prostate cancer often have a greater fraction of cPSA and a lower percentage of fPSA compared to men without the disease [2, p. 2752].

This has led to the measurement %fPSA, the percentage of the ratio of free to total PSA. This is particularly useful for men in the diagnostic grey zone, with total PSA levels between 4 and 10 ng/ml, where total PSA is not helpful as a stratification tool. Currently, the U. S. Food and Drug Administration (FDA) has approved %fPSA to aid PSA testing in men with a normal DRE and small PSA elevations within this diagnostic grey zone [2, p. 2753].

The effect of PSA on incidence and mortality

The recorded incidence of prostate cancer peaked in 1992, a few years after the introduction of PSA. Both the number of prostate cancer cases and the mortality of the disease increased rapidly but has declined since then. The mortality and incidence rates from 1986 to 2013, and an estimation of the rates from 2013 to 2016, can be seen in Figure 2.4 and 2.5.

One explanation of the increased incidence rate after the introduction of PSA is the large amount of prostate cancers that were unknown and had not yet been detected without the use of PSA. The number of diagnoses each year has decreased again but there is a difference in the aggressiveness of the cancers that are detected compared to before since a higher proportion of the less aggressive cancers are now found.

The mortality rate is lower now than before, but the decline in prostate cancer-related deaths can not only be explained by the use of PSA. In addition to PSA screening, better biopsy methods and increased public awareness about prostate cancer has benefited the earlier detection and decrease in mortality [2, p. 2735].

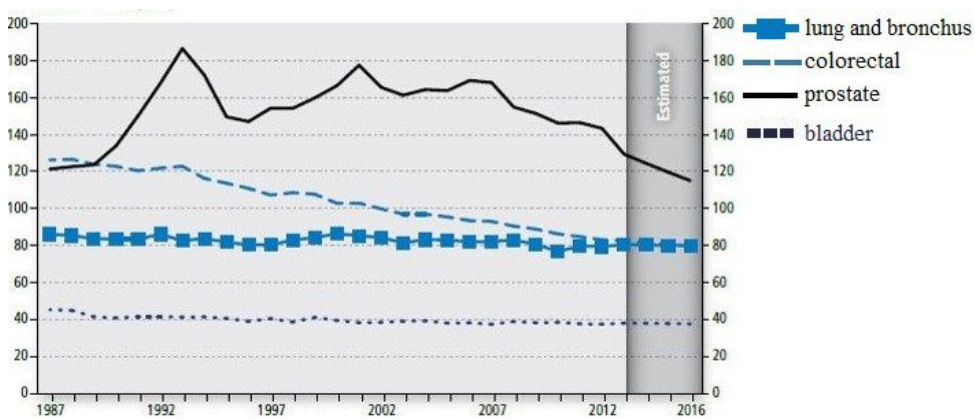


Figure 2.4: Incidence rate of a number of common cancers in men in Canada, prostate cancer incidence visualized as a black line [13].

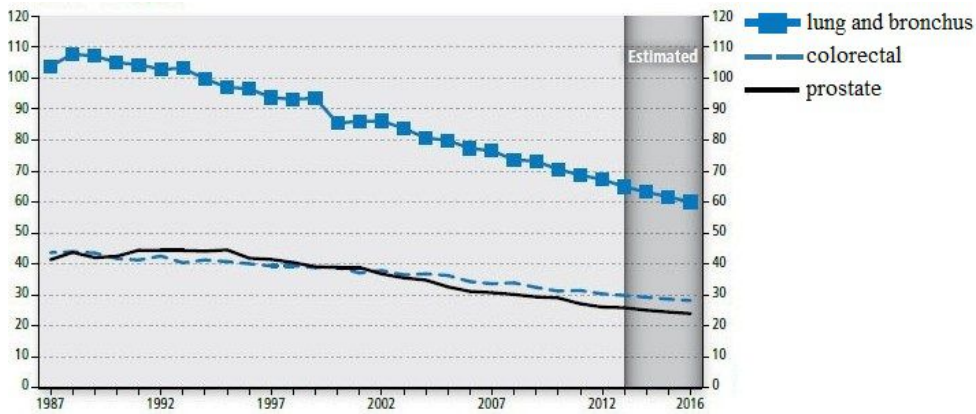


Figure 2.5: Mortality rate of a number of common cancers in Canada, prostate cancer mortality visualized as a black line [13].

As mentioned earlier, two randomized trials have assessed the effect of PSA screening on prostate cancer mortality with controversial results. The PLCO cancer screening trial reported no difference in prostate cancer mortality after 7 years of follow up between the group that was receiving annual screening and usual care. On the other hand, the ERSPC reported a reduction in the rate of death from prostate cancer by 20% after a median follow-up of 9 years but after a large amount of biopsies that might have been unnecessary. They also reported reductions in aggressive cancer cases. These studies show that prostate cancer has a very low disease-specific mortality and will only impact life expectancy in relatively few men. However, both trials reported results after a short time considering the long natural history of prostate cancer and the results might change with time [2, p. 2763].

Introducing PSA mainly affected the number of detected low-grade cancers and the mortality effect due to this will take longer time before it affects the data [2, p. 2704]. Also, the suffering and side effects from screening and treatment need to be evaluated to conclude the need for better screening methods. More large scale clinical studies are needed to determine the effects.

In contrast to the mortality rate, the migration to a higher rate of low grade cancers can be traced back to PSA [2, p. 2704]. Since the initiation of PSA screening, the incidence of local disease has increased while the incidence of metastatic disease has decreased. Many of the low-grade diseases that are found are in younger men, the incidence of prostate cancer in men 50 to 59 years of age has increased by 50% after the introduction of PSA screening [2, p. 2771]. Today, approximately 90% of the detected cases are in a clinically localized stage where the disease is not yet metastatic [2, p. 2771]. This has

implications on the need for, type of, and complications after therapy which need to be addressed.

2.2.5 Treatment

Detected prostate cancers are graded with a Gleason score. Depending on the stratification into low or intermediate/high risk, treatment is decided. The treatment also depends on criteria such as age and prognosis, since different risks and benefits are associated with every treatment. Available treatments today include Radical Prostatectomy (RP), brachytherapy, Intensity-Modulated Radiation Therapy (IMRT) and Androgen Deprivation Therapy (ADT). The goal is to cure the disease, while keeping the costs down and maintaining a high quality of life for the patient, for example by maintaining urinary continence and avoiding erectile dysfunction [1].

The most common treatment is RP, where the prostate and some surrounding tissue is surgically removed. It was the first available treatment for the disease and has been in use for more than 100 years [2, p. 2775]. Brachytherapy is a treatment where a radioactive substances is introduced into or close to the tumours. This treatment has been in use for many years but the method has been refined a lot and is therefore increasingly used [2, p. 2864]. IMRT is a type of External Beam Radiation Therapy (EBRT) where a large number of radiation beams with varying intensities are delivered at the cancer tumour [14]. ADT is one of the most effective treatments for tumours and consists of suppression of testosterone which causes the cancer cells to shrink or grow more slowly. However, the main disadvantage of this type of treatment is that it often leads to hormone resistance, Castration-Resistant Prostate Cancer (CRPC), which is an aggressive and very serious condition which most likely will lead to death [2, p. 2954].

Traditionally, patients with a low risk cancer who have a relatively long life expectancy are referred to Active Surveillance (AS), which is almost unique to prostate cancer [2, p. 2772]. AS is a management strategy that is used to delay or avoid the morbidities associated with treatment. The patients are monitored closely for any growth or other differences of the tumour but no active treatment is started until the cancer has progressed enough for active treatment to be considered. The patients in AS are traditionally examined with regular PSA measurements, clinical examinations and prostate biopsies [15]. The risk that has to be considered is that some cancers will progress during AS from a curable disease to a potentially incurable one. However, in many studies, only 25 to 50% of patients in AS had any tumor progression within 5 years and less than 5% of the patients had a disease that became incurable [2, p. 2772].

2.3 New testing methods

The implementation of one or several new tests in the screening and diagnosis of prostate cancer has the potential of reducing the number of performed unnecessary biopsies. This may increase the cost-effectiveness for the health care system and the patients from screening until end of life.

A schematic diagram over the whole process from screening to treatment, can be seen in Figure 2.6 [11]. The numbers 1–4 represent phases that are currently not in use but where new tests could potentially be introduced to reduce costs or increase quality of life for the patients. Number 1 represents the initiation of new tests in the screening phase, number 2 after a positive biopsy, number 3 after a negative biopsy and number 4 in the treatment decision. The screening phase, 1, and the phase after a negative biopsy, 3, are those considered in this project. The biopsy method considered together with the new tests is TRUSGB.

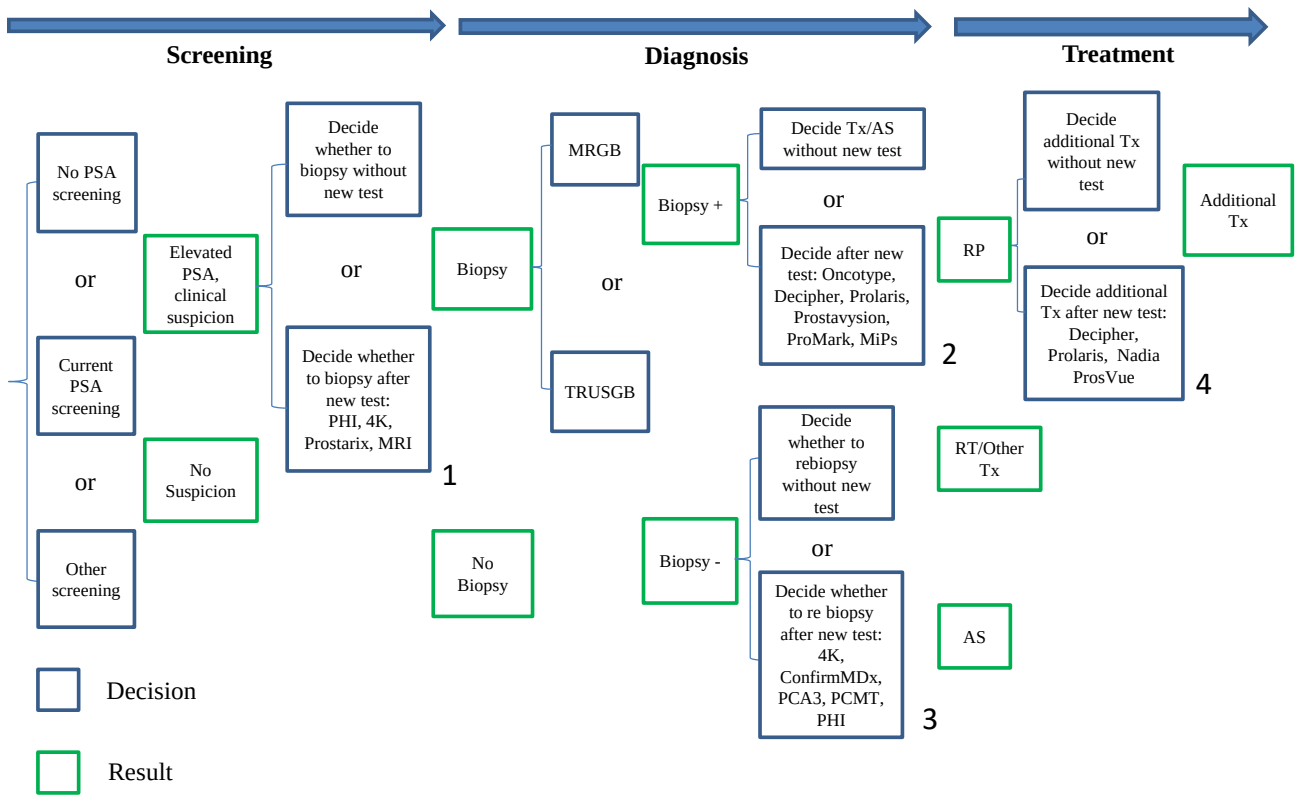


Figure 2.6: Flow chart of prostate cancer phases, developed by the local team. MRGB = Magnetic Resonance imaging-Guided Biopsy, TRUSGB = Transrectal Ultrasound-Guided Biopsy, Tx = treatment, RP = Radical Prostatectomy, RT = Radiation Therapy, AS = Active Surveillance. The numbers 1–4 represent potential phases to introduce new methods. The phases considered in this project are phase 1 and 3 which are the screening phase and the phase after a negative biopsy respectively.

2.3.1 Screening phase

The goal in the initial screening phase is early detection of patients with significant prostate cancer among a healthy population. The ideal screening test would detect all patients with significant cancer and exclude all patients with no risk of having the disease, for example patients with a benign enlargement of the prostate. A new test in this phase would help the clinicians to stratify the patients that are in need of a prostate biopsy to diagnose an eventual cancer.

Four new tests for use in the screening phase have been assessed to investigate the cost-effectiveness of implementing them in clinical practice: Prostarix, 4Kscore®, MRI and Prostate Health Index (PHI).

Prostarix

Prostarix is a urine test performed after a DRE. The DRE disturbs the prostate and releases substances into the urinary tract, which can be measured in the first catch urine. Several metabolites (sarcosine, alanine, glycine and glutamate) that have been associated with prostate cancer are measured from serum samples to generate a risk score through a logistic regression algorithm [12]. The test result correlates with the aggressiveness of the cancer but is not affected by normal enlargement of the prostate [16].

4Kscore®

The 4Kscore®, also called the four-kallikrein panel, is a value of the risk of finding aggressive prostate cancer in percent [17]. A kallikrein is a subgroup of enzymes cleaving protein peptide bonds. The score is calculated by an algorithm with the value of bio-marker measured from a blood sample combined with other clinical information about the patient such as the patient's age, results from prior biopsies and the assessment from a DRE [18]. Four bio-markers are measured in the blood sample: total PSA, fPSA, intact PSA and human Kallikrein-2 (hK2) [18]. The active forms of PSA and hK2 are produced in the prostate and are released in the prostate fluid which is found in small amounts in the blood [17].

The threshold for when to perform a biopsy after a 4Kscore® test can be customized for each patient depending on the health state and how risk-averse the patient is [17]. For example, an old man with low life expectancy might have a higher threshold for when a biopsy is considered than a younger healthier man.

Magnetic Resonance Imaging

MRI is one of the most common imaging methods. It shows the soft tissues in the human body and metabolic processes by using the interaction of nuclear spin and an external magnetic field to build images [19]. The non-invasive method can be used to detect, monitor or measure prostate cancer. Tumors can be seen in the image due to their high metabolism [20]. If a biopsy is performed after the scan, less biopsy samples might be needed due to the knowledge of the tumour and the location of it. This may lead to fewer complications for the patients and a more reliable diagnosis [6].

Prostate Health Index

The PHI test is blood based, measuring three bio-markers which provide information about the cause of an elevated PSA [21]. The PHI is obtained by an algorithm including PSA, fPSA and [-2]proPSA (p2PSA) [21]. The bio-marker p2PSA is an inactive enzyme and an isoform of PSA which is correlating with the aggressiveness of the cancer [22].

The test is recommended by FDA to be used in patients aged 50 years or older with a PSA between 4 and 10 ng/mL. It was approved by FDA in June 2012 and can be used to indicate a risk of prostate cancer. There is no specific cut-off approved by FDA but they state that physicians may recommend biopsy for patients with a PHI result over 55 but not for a patient with a PHI result under 27 [23].

2.3.2 After a negative biopsy

In current practice, a high proportion of the initial biopsies with negative results are missed cancers. A biopsy can never provide a guarantee that no cancer is present in any part of the prostate, even if repeated. Because of this, many are forced to undergo regular repeat biopsies due to the risk of having the disease. A new test used after a negative initial biopsy would help the clinicians to predict the second biopsy result, leading to a better patient stratification.

Apart from PHI and 4Kscore®, which can be used both in the screening phase and after a negative biopsy, three more methods have been reviewed: Prostate Cancer gen 3 (PCA3), Prostate Core Mitomic Test (PCMT) and ConfirmMDx.

Prostate Cancer gen 3

PCA3 is a prostate-specific gene that is expressed 60 to 100 times more in prostate cancer cells than in normal cells. The gene is not affected by normal prostate enlargement or other benign prostate conditions [2, p. 2758]. The test has shown utility in finding even very small cancers within a large background of normal cells [2, p. 2758].

The PCA3 assay is a urine test performed after a DRE. The score is calculated by the result from two nucleic acid amplification tests. One test is used for detection of PCA3 mRNA which is a non-coding RNA expressed in prostate cancer tissue, the other one for detection of PSA mRNA which is relatively constant in normal prostate cells [7]. The ratio between PCA3 mRNA and PSA mRNA is the resulting PCA3 score [24].

The PCA3 assay is recommended by FDA to aid in the decision for repeat biopsy in men 50 years of age or older using a cut-off of 25, where a PCA3 score less than 25 is associated with a decreased likelihood of a positive repeat biopsy [24]. It is indicated for use on men with one or more previous negative repeat biopsies who would be recommended a repeat biopsy by a urologist [25].

Prostate Core Mitomic Test

PCMT is a tissue-based molecular test measuring a deletion of mitochondrial DNA. This DNA is circular and consists of approximately 16,500 base pairs, coding for 37 genes. The test is performed on tissue from a previous biopsy and Reverse Transcription Polymerase Chain Reaction (RT-PCR) is used to detect the DNA deletions [26]. The deletions looked for are the removal of a number of mitochondrial-encoded genes required for the electron transport chain. It appears in and around prostate cancer cells but is absent in normal tissue, creating a cancerization field effect that can not be seen in a microscope. When cancer is present, this field effect is extending through the prostate which makes the test highly specific [27].

ConfirmMDx

ConfirmMDx is a tissue-based genomic test, just as the PCMT. The assay involves Quantitative Methylation Specific Polymerase Chain Reaction (PCR) detecting the cancerization field based on methylation of DNA, specifically the cancer associated bio markers GSTP1, APC and RASSF1 [26].

2.4 Systematic Literature Review

A systematic literature review requires considerably more effort than a traditional review, where the search is not as extensive. The aim of a systematic literature review is to identify all available evidence from previous studies relevant to a particular question. The quality of the studies is assessed and information is extracted from the studies of acceptable quality that meet the predetermined inclusion and exclusion criteria of the review [5, p. 221].

Compared to a traditional review, the statistical power in the results and the ability to study the consistency of results are improved. All studies on the same topic are combined, the assessment of study quality is standardized and the data synthesis of studies is more systematic. Furthermore, similar data collected from several studies show evidence of robustness and transferability. If the data is not consistent through the studies, this variability can be investigated and discussed. This is particularly important for the most sensitive elements in the study [5, pp. 221–222].

The improved statistical power from the systematic review can also be a disadvantage. The systematic review will detect small effects from the studies, but also small biases will be reflected in the results. This is one main reason to carefully assess the quality of every individual study. Using more than one reviewer can also increase consistency and accuracy of the assessments [5, pp. 221–222].

The process of a systematic literature review can be divided into five main steps [5, pp. 221–222]:

1. **List questions.** It is helpful to start with a table of variables that are sought for in the review and their corresponding definitions. It is important to specify the components of the questions before starting to ensure high quality of the result.

2. **Find available studies.** The aim in a systematic review is to find all studies on a question. This is hard to achieve in practice as there are a high number of published studies available on several databases. Search algorithms with a smart design can simplify this process.

3. **Select and assess studies.** The quality of all identified studies should be assessed with a standardized approach. The conclusions from the review should be based on high quality data. Presenting the quality highlights the strength of evidence for any recommendations made.

4. **Summarize and synthesize collected data.** Relevant data from selected studies is extracted and summarized.

5. **Validate results.** Applicability of the results and conclusions are validated by examining for example the difference in settings, methods and results between the studies.

In a systematic review, the objectives of the analyzed studies are important. Studies assessing the same intervention might have different aims and approaches. For example, some research the performance of the test and other the outcome after introducing a test on the market. In this project, only clinical utility studies were included.

When introducing new interventions to the market, clinical utility of the tests often needs to be proven, but there are different definitions of the term. There is a lack of clarity of how to measure and judge the clinical utility and there is not a formal definition used by all. The word utility literally means usefulness, but the question is how to define usefulness and how to measure it [28]. Clinical utility can include several variables, but mainly judgments about benefits and drawbacks together with the usefulness of the intervention [28].

In this project, the clinical utility is assessed as the risks and benefits of using the tests clinically and the ability of the tests to affect outcomes and decisions compared to not using them.

2.5 Modeling

Models can be used to predict health economical outcomes and the effect on people in a certain scenario or after a particular intervention. The scenarios can for example be the progress of a disease in a population over a specific time period and outcomes may be costs, mortality or quality of life. A model can provide a framework for decision making so that analysis can be made in a systematic way. They simplify complex scenarios to something manageable and can be visualized and analyzed in computer programs, for example TreeAge Pro 2017 [29] which is used in this project.

In several studies only a couple of years are covered, especially when evaluating new interventions. Models can be used to predict future events and calculate the cost-effectiveness even for novel interventions.

Health economical models are usually developed to evaluate a certain intervention or scenario over a longer period of time. Calculations are performed to predict the value today of introducing a new intervention. Future costs have less impact today than in the future, due to the interest rate. The same applies to health consequences, but is a bit more controversial. The differences are calculated by adding a discount rate to compute future val-

ues (FV) of costs and health outcomes in the model [5, p. 272]. The general equation for calculating the present value (PV) is:

$$PV = \frac{FV}{(1+r)^t}.$$

where r is the discount rate as and t is the time. The discount rate most commonly used for cost-effectiveness analyses is 5% (r = 0.05) [5, p. 272].

2.6 Background of health economics

Health care decisions are complex and tools are needed for evaluation of different options. The choices may have a large impact on costs and quality of life for patients. Each decision, for example which treatment to use, has uncertainties and trade-offs that need to be accounted for but the decisions are often too complex for the human brain to comprehend. The goal is that the benefits are greater than the disadvantages in every decision [5, p. 3].

The decision makers need guidelines to determine the most beneficial alternative in each situation. To conduct these guidelines, decision analysis can be performed by calculating costs and utilities for the alternatives. Some uncertainties will always be present but by combining uncertainties and trade-offs the comparison gets more conceivable. The most beneficial option has the best balance between benefits and risks [5, p. 5]. However, it is not possible to compare the outcomes and trade-offs straight off.

Calculation tools such as cost-effectiveness analyses can be used to help the decision makers analyze options. The method compares the costs related to a certain decision with the health effects. The health effectiveness can be measured in, for example cured patients, prevented diseases or saved lives. Other measures for the effect are preference based, for example Quality-Adjusted Life Years (QALY) gained [5, p. 267].

QALY includes length of life and reflects how an individual values the quality of life. It represents how limited a patient feels by a certain disease or disability. QALY is calculated by multiplying the utility value associated with a certain health state with the number of years lived in it. The utility is measured on a scale from 0 to 1, where 0 is death and 1 is perfect health. It can be measured by utility derivations such as time trade-offs or by physiological scaling where patients rate their own health. Time trade-offs show how many years of life a person with a certain condition would trade for perfect health. Utility based measures are the most commonly used in QALY calculations, since there is no proof that the numerical values from physiological scaling would represent the actual preferences [5, pp. 97–104].

After calculating cost and effectiveness for each alternative the measurements can be compared. The assessed options are compared to a base case which is selected from the beginning. Options with higher cost and lower effectiveness than the base case are dominated and can be excluded while the dominating options can be evaluated further.

2.6.1 Types of health economical models

Several types of models are used in health economics, but the most common ones are decision trees and Markov models. A decision tree is a simple way to model a decision problem. It structures the alternatives and gives a visual representation of all possible options and the consequences that may follow [5, p. 80]. The model begins with one branch which divides in case of events such as disease progression, as in Figure 2.7. A probability is assigned to every option, as can be seen in the example in Figure 2.8. The model is visually intuitive but has limitations. It can not reflect scenarios that are time dependent or handle repeated events which usually are optimal for health economic scenarios [30].

When developing decision models, nodes are used to visualize transitions between the events and states. The nodes used in the model in this project are the following, which can also be seen in Figure 2.7:

- **Chance Nodes.** Represent possible outcomes of an event based on probabilities.
- **Terminal Nodes.** Located at the end of each branch, terminating every cycle. In combination with the terminal node, the state in which the individuals will begin the next cycle is presented.
- **Decision Nodes.** Represent a decision made by a decision maker, for example a clinician. Only one of the available actions can be chosen and the outcome only depends on the decision, not probabilities. The different decisions are compared when running the model.
- **Markov Nodes.** Mark the start of a Markov cycle tree. All Markov states originate from this node, as well as the modeled patients at the start of every cycle.

A decision tree does not have the capacity to model events that are occurring and recurring over a period of time, when risk over time is involved or when the timing of events is important without making unrealistic assumptions [31]. In that case, Markov models are a better alternative. The events

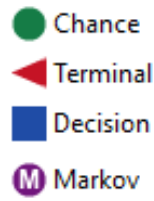


Figure 2.7: Visualization of the nodes in the TreeAge model [29].

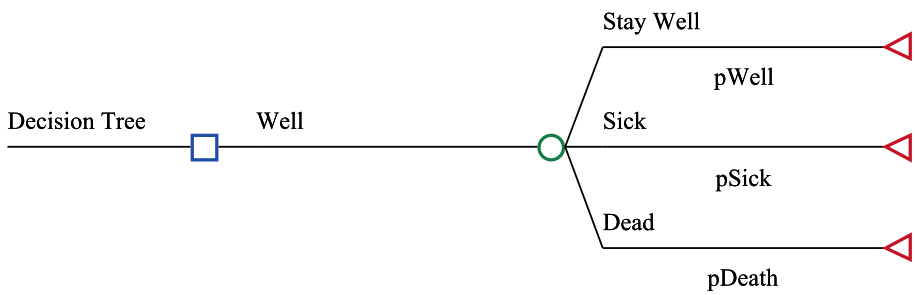


Figure 2.8: Simple example of a decision tree where a patient is well and proceeds to staying well, getting sick or dying.

in Markov models are modeled as transitions between health states and the model takes into consideration events from previous cycles [5, p. 305]. The model is run for a predetermined time horizon. Ideally, it is run until all patients are in the dead state but usually over a set of time spans such as 5, 10 or 20 years. The time horizon is divided into equal time periods where one period is called a cycle [5, pp. 307–309]. The length of the cycles is determined by the data in the model and depends on the modeled scenario, but a common cycle length is one year.

The Markov model can be visualized as a state transition diagram where the health states are represented by circles. Arrows show possible transitions between the health states [5, p. 305], as in the example in Figure 2.9. Patients will always be in one of the health states, which are assigned with a specific utility [31]. The incremental utility of every patient depends on the states the patient has been in and the time spent in them [31].

Each transition between the states depends on events. The events are visualized in Markov-cycle trees, as in Figure 2.10, which have visual similarities with decision trees but can be run in cycles [31]. Each cycle begins at the Markov node and the group of patient pass through the tree, from this node to a terminal node, in each cycle [5, p. 329]. In the following cycle the population starts from the Markov node again. The population is divided

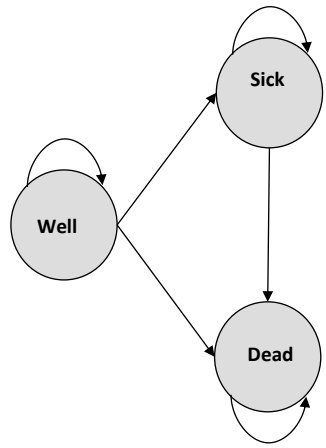


Figure 2.9: Example of a simple state transition diagram, with the health states Well, Sick and Dead.

into the states depending on which state they ended previous cycle in. Some states allow the patients to remain over several cycles while other states are temporary and thus have transitions to other states but not to itself. This can be useful for example in situations when there is a one time cost, such as performance of an intervention [5, pp. 323–324].

2.6.2 Evaluation of health economical models

There are three basic methods to evaluate a Markov model: the fundamental matrix solution, cohort simulation and Monte Carlo simulations [5, pp. 311–312]. In this project, Monte Carlo simulations were used for evaluation. In a Monte Carlo simulation, an amount of individual patients are simulated to pass through the Markov model. The starting point of the patients needs to be set initially, after that they will proceed from cycle to cycle based on the transition probabilities. The patients take different paths through the model until they reach terminal nodes, and depending on that begin the next cycle in a certain state [5, p. 320].

The models normally contain an absorbing state, such as death, and the simulation is run until all of the patients have reached this or for a predetermined number of cycles. The definition of an absorbing state is that the patients will remain in that state for all following cycles after entering it, there are no possible transitions from the state except back to itself. The number of cycles spent in every state is finally recorded for every patient [5, p. 326].

One benefit with the Monte Carlo simulation is that the variation around

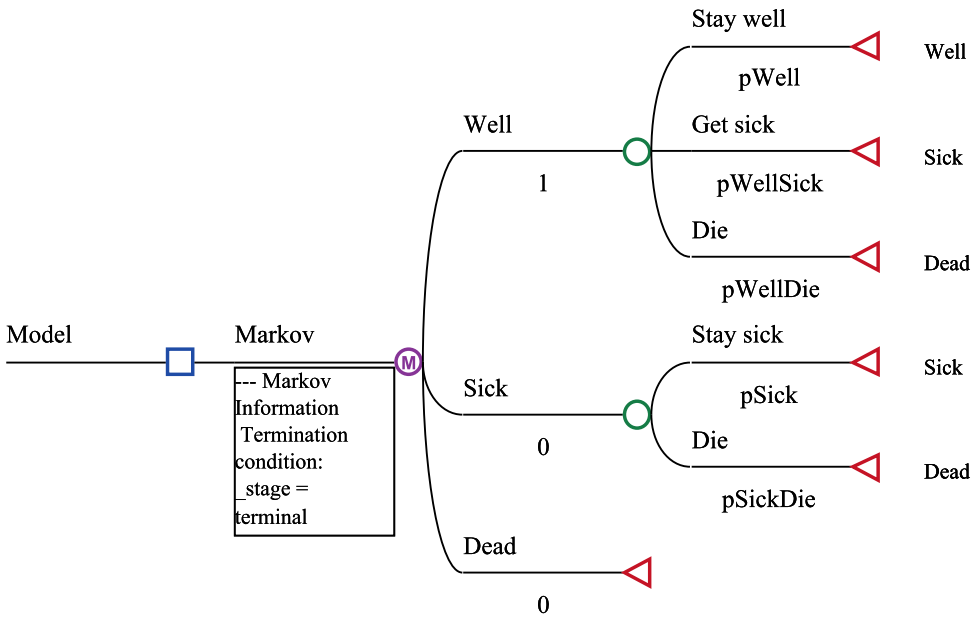


Figure 2.10: Example of a Markov model, with the health states Well, Sick and Dead. The variables under the branches are transition probabilities to every branch and it can be seen that all patients begin in the Well state in the first cycle.

the mean can be obtained since the simulation yields data for every individual patient. Another benefit is that the probabilities can be defined as functions that depend on data of previous events, such as the order that the events happened in or the time between the events. To keep track of previous events, variables can be assigned to the patients depending on the states they pass through in a cycle. The variables are reset for every patient and are in this context called tracker variables [5, pp. 320–323].

Monte Carlo microsimulations

The type of Monte Carlo simulations used in this project were Monte Carlo microsimulations, simulating one individual at a time [5, p. 345]. Attributes can then be assigned to the individuals instead of modeling them as states [30]. In this way it is possible to follow every individual in the model and track each patient's disease history [5, p. 347].

The Monte Carlo microsimulation tracks the individual until a specific termination criteria. A disadvantage is that it can be very time consuming since every data point needs one simulation and usually simulations for a large number of individuals are needed to get reliable results [30].

2.7 Sensitivity Analysis

The collected data used in the model is sometimes established in literature, but in other cases the numbers are not as definite. A model is a simplification of the reality and assumptions need to be made to reduce the complexity. Every assumption introduces uncertainties to the model.

To understand the impact of changes in these numbers and assumptions a sensitivity analysis can be performed [5, pp. 239–241]. This is done by varying uncertain variables in all reasonable ranges to understand the effect on the results. If a small change of the values gives a big change in the results, it implies that the precise values of these variables need to be researched further. If the result is not sensitive to changes of the variables, the precise values of the parameters are irrelevant [32]. The variables that can be analyzed include the time horizon, transition probabilities, costs and utilities for the various states [5, p. 334].

A sensitivity analysis can be deterministic or probabilistic. In a deterministic sensitivity analysis, the variables are altered one at a time. The probabilistic analysis includes second order uncertainties by varying several variables together [32].

3 Methodology

3.1 Systematic Literature Review

A systematic literature review was conducted to assess new tests that could be used clinically in the initial screening phase or the diagnostic phase, after a negative biopsy, of prostate cancer. The review was performed to systematically go through a larger number of articles and extract as much relevant data as possible. The researched screening tests include MRI, PHI, 4Kscore® and Prostarix. The researched tests for use after a negative biopsy were PHI, 4Kscore®, ConfirmMDx, PCA3 and PCMT. These tests were considered as the methods with highest evidence of clinical outcomes which are most likely to be adopted in clinical practice.

The systematic literature review was performed in collaboration with G. Olleik and A. Tarifi. All the steps of the review were performed by two different persons separately to eliminate errors and biases in the final results. The phases were divided, E. Bonnevier was responsible for phase 1 and E. Palenius for phase 3. Phase 2 and 4 were assessed by other students in the local team but not as a part of this project. The work in the systematic literature review was continuously co-assessed by G. Olleik and differences were discussed before making final decisions in every step to ensure high quality. The purpose of discussing the articles was to ensure quality, learn from each other and to speed up the process. The systematic literature review was organized by the local team in a predetermined number of stages: selection of studies, data extraction, assessment of quality and data synthesis.

The review allowed evaluation of all available studies on the subject to identify information about the clinical utility of the tests. The studies were extracted from the databases Cochrane [Wiley], Embase [Ovid], Medline [Ovid] and Web of Science [Thomson Reuters] with help from Ibtisam Mahmoud and Elena Guadagno, who are librarians at the MUHC medical library. All articles were managed in the reference managing software Endnote [33].

The desired outcome from the systematic literature review was information that could be converted into numbers to use in the models. Specifically,

researched numbers included measures of the proportion of patients with a change in decision to perform a first or repeat biopsy due to the new tests and the implications of this in following cycles. The number of unnecessary prostate biopsies and missed cancer diagnoses after introduction of the tests as well as the direct cost in the Quebec health care system perspective for every test was sought for.

3.1.1 Selection of studies

To ensure consistency for all researched tests, inclusion and exclusion criteria were formulated based on the desired outcome. The criteria were used in every step of the systematic review. Tests with no included articles to assess were excluded. The inclusion and exclusion criteria were:

Inclusion criteria:

- Studies related to screening, diagnosis or treatment of Prostate Cancer
- Studies related to the assessment of the test
- Clinical utility studies regarding the test

Exclusion criteria:

- Conference abstracts
- Unrelated articles
- Untraceable articles
- Commentaries

During the initial phase of the review, titles and abstracts of all articles were screened to quickly exclude irrelevant studies. The articles that were included after this step were screened in full text before proceeding to data extraction with all relevant included articles.

3.1.2 Data extraction

A data extraction form was formulated with the local team to ensure consistency and precision in the data extraction. The form was based on a data extraction template from a study by S. Sommariva et al. [34]. It includes questions on the study characteristics, number of participants, outcomes measured, what phase the test was performed in, risk for bias, funding resources and extracted results. The resulting form can be seen in Appendix A.

The included articles were assessed in detail and the relevant data was manually extracted into the data extraction form. Quantitative numbers that were relevant for this project were extracted and recalculated when needed, for example into probabilities when given as rates.

3.1.3 Quality assessment

A quality assessment form was created in collaboration with the local team to assess the quality of the studies, based on a score from 0 to 2 on a number of criteria. The quality assessment form was created based on the Rector et al. checklist [35] and includes factors such as bias, financial interests and study design. The resulting form can be seen in Appendix B.

Each criteria was scored from 0 to 2, or set to Not Applicable (N/A), for every article. A final score was calculated as the percentage of the total score, excluding the questions set to N/A. The articles were categorized in excellent quality (score >0.75), good quality (score 0.50–0.75) or poor quality (score <0.50). The quality was assessed to ensure high quality in the studies used for data extraction. An article with high quality ideally has no biases, clear descriptions of what was done and close resemblance to clinical practice.

3.1.4 Data synthesis

When the data extraction and quality assessment was performed and discussed for every article, the relevant results were inserted into an excel data sheet. This was done to get an overview of the available data from all studies in every test. The data was compared between the studies to research if the results were coherent. By creating one data extraction excel sheet for every test, it was possible to compare the amount of extracted relevant data and the number of relevant references. If the variations in the extracted data were large or if there was too little available data in this step, the test was excluded. These tests were not modeled due to the large uncertainties.

A set of data points was chosen for the modeling of each included test. The data was selected from the studies with the highest rated quality that included relevant data. The articles were set as the primary references. When applicable, a confirmatory reference was also chosen to confirm the data. The selected values were set as the base case for each test, which was later altered and evaluated in a sensitivity analysis.

The costs for the new assessed tests were not found during the systematic literature review. Therefore, a small review was conducted and the companies developing the tests were contacted.

3.2 Modeling

An existing Markov model previously created and published by members from a team at the Research Institute of the MUHC [6] was reviewed in the initial phase of the project to get an overview of the project and the work that was previously done. In the previous study, the cost-effectiveness of the current practice using biopsy with ultrasound, TRUSGB, was modeled as well as the implementation of MRI in combination with MRGB. The results from the previous study were used to compare with the results for the new tests. The developed models were based on the structure of the existing model for comparability.

The existing TRUSGB model from the previous study can be seen in Appendix E [6]. In that model, current practice was modeled from the screening phase due to elevated risk for prostate cancer until the end of the patient's life. The patients were initially stratified according to the initial biopsy result and were either diagnosed with prostate cancer or referred to Follow Up (FU) where the patients are continuously examined with regular biopsies and other examinations. The patients were then followed through diagnoses and treatments and the incremental costs and effectivenesses were recorded. A summary of the results from this previous study are presented in Appendix G and H.

3.2.1 Development of Markov model

The TRUSGB model from [6] was modified to fit the phases studied in this project. The goal was to make similar assumptions in the new models as in the previous model to be able to compare the results. One model was created for each evaluated test. Only the tests with enough resulting data from the systematic literature review were modeled. The variables depending on the initiation of the new tests were altered and remaining values was used from the previous model. The models were built in the modeling software TreeAge Pro 2017 as Markov cycle trees. Markov modeling was chosen due to comparability to the previous models, as well as their usefulness in health care scenarios.

The extracted data from the literature review was inserted into the models. In cases where variables were not extracted during the systematic literature review, values from the previous study were used. In applicable cases, these variables were researched in the sensitivity analysis. All utilities used in the model were also extracted from [6] and similarly assessed in the sensitivity analysis.

Screening model

A base model for the screening tests was developed based on the previous model. The modeled population was men referred to an initial biopsy in the screening process. Screening tests could make a second stratification for these patients and thus reduce the number of unnecessary biopsies.

In the previous model, the initial stratification of patients depended on the result of the initial biopsy. The main change in the screening model was the addition of a screening test instead of going straight to biopsy. In the developed screening model, the patients were stratified into initial biopsy in case of a positive test result or to FU in case of a negative test result. The patients with a negative biopsy result were also referred to FU while the others proceeded in the model with a prostate cancer diagnosis.

Model after a negative biopsy

A base model for the tests after a negative biopsy was developed. One state was added in the TreeAge model, for patients performing a repeat biopsy. In the previous model, all patients with a negative initial biopsy were referred to FU where the eventual repeat biopsies took place.

In the model created in this project, the patients with a negative initial biopsy were first stratified depending on their need for an immediate repeat biopsy. In the repeat biopsy state, the new test is performed before another stratification is made according to the test result. The patients with a negative test result or a negative second biopsy then go to FU while a number of patients get a positive biopsy and therefore start the next cycle in the *Positive Biopsy* state.

3.2.2 Evaluation of Markov model

The models were run separately in TreeAge using Monte Carlo microsimulations for 10,000 individual patients. Each model was run 100 separate times and the mean cost and QALY was registered. All analyses were performed for 5, 10, 15 and 20 year time frames to be able to compare the differences in cost and QALY between the tests in different periods of time.

To determine the dominating strategy, the confidence intervals of the costs and QALY for every test were compared to the results for TRUSGB and MRGB from the previous study. If the costs were lower and the QALY were higher the strategy were considered as dominant.

3.3 Sensitivity analysis

A deterministic sensitivity analysis was performed, where one value was changed at a time in predetermined ranges to determine the impact on the results. The analysis was conducted in TreeAge by running the models with the chosen values. The model was run for 10000 individuals and each model was run 10 times.

The sensitivity analysis was performed to evaluate the data in the models and compare different scenarios. Data was collected from studies with good quality that had confirmatory references but there are uncertainties. For example, costs change and are not equivalent in different countries. Also, the studies have limitations which incorporate uncertainties in the data. There might be bias in the results or assumptions that were made in the study design. The collected data was reviewed to identify the main uncertainties. The data was classified as uncertain in case of few references or if the values differed considerably between the sources.

The same values that were altered in the sensitivity analysis in the previous study were included in the analysis for comparison, see Appendix G. These variables included the discount rate, rate of patients with low risk prostate cancer referred to AS, probability of recurrence among patients with intermediate/high risk prostate cancer and utilities for all states. Two AS rates (40% and 50%) were added in the sensitivity analyses for the new tests as well as MRGB. Apart from the analyzed utilities in the previous study, utility 1 for all states was added in all sensitivity analyses.

The FU cost was added to the sensitivity analyses of all tests. In the TRUSGB model, the patients in FU are expected to undergo biopsies every 3–4 years. After implementation of the new test, there will probably be fewer repeat biopsies due to a better initial stratification and thus this cost is expected to decrease. The scenarios of biopsies every fourth or fifth year are investigated in the sensitivity analysis, as well as every second year as a comparison. In addition, one repeat biopsy is already modeled in the phase after a negative biopsy.

Apart from the variables included in the sensitivity analyses for all tests, a number of additional variables were altered for every test. These values depended on the available data from the systematic review for each test and what values had the most uncertainties.

3.3.1 Screening tests

4Kscore®

Two variables were added in the sensitivity analysis for 4Kscore®. These were the probability of an initial biopsy after performing the test and the probability of a positive initial biopsy result.

The reason for including the probability of an initial biopsy is that numbers were presented in a wide range. The values that were chosen in the sensitivity analysis were the lowest (40%, [36]) and highest (63.8%, [37]) numbers presented in the articles compared to 48.7% [38] which was used in the base case.

The reason for including the probability of a positive biopsy result is that this number was expected to increase after implementation of the test since many of the biopsies that would have been negative were not performed. This was not the case in the studies from the systematic literature review, and thus it was tested in the sensitivity analysis instead. The values that were chosen in the analysis were the lowest probability (33%, [37]) presented in the articles as well as the probability from the previous study (56.5%, [6]). These were used together with 65% which was a 50% increase compared to the probability in the base model (43.3% [38]).

PHI

Three variables were added in the sensitivity analysis for PHI: the cost of the test, the probability of an initial biopsy after performing the test and the cut-off used by the clinicians to stratify the patients.

The reason for including a variation of the test cost in the analysis is that the references presented costs with a big variation, probably due to geographical differences. The cost in the model ($150, [39]) was extracted from a European source and the chosen values were from an American source ($670, [40]) and a mean of the two ($410).

The probability of an initial biopsy after performing a PHI test was included because of the same reason as in the sensitivity analysis for 4Kscore®. The values were the highest presented value in the articles (84.5%, [41]) as well as a mean (66.25%) of the highest value and the value for the base case model (48%, [42]).

In the FU state, two groups of people were combined. These were the patients that did not get a biopsy because of a negative PHI and the ones who had a negative initial biopsy. The first group includes men with missed cancers due to the test and the other group due to the biopsy. The false

negative rate for the missed cancers after the test was used in the model. The values that were assessed in the sensitivity analysis were the false negative rate for biopsy (10.4% [6]) and a mean of this and the false negative rate for PHI (19.8%).

There is no recommended cut-off for the PHI test, therefore one was chosen for the base case and the others included in the sensitivity analysis. 41.5 [42] was chosen for the base case since it was presented as the best combination between sensitivity and specificity. FDA has approved PHI with a recommendation to not biopsy with a cut-off under 27 definitely biopsy with a cut-off over 55 [23]. This is a wide range and therefor different cut-offs was studied. Other cut-offs analyzed were 27.6 [42], 40.3 [41] and 50.9 [41].

3.3.2 Tests after a negative biopsy

PCA3

The additional variable that was analyzed in the sensitivity analysis for PCA3 was the probability of a repeat biopsy after performing the test. The rates that were assessed in the sensitivity analysis were 37% [43], which is the corresponding value in the confirmatory reference study, and 60%, which is 20% higher than the chosen base case value (50.4%, [44]). The stratification of patients to perform a repeat biopsy is the main difference of using the test and a 20% higher rate was assessed to see the impact on and make sure not to underestimate the cost.

ConfirmMDx

The variable that was added in the sensitivity analysis for ConfirmMDx was the test cost. The cost in the model ($4,440, [40]) originated from a source from the United States, where health care costs are often higher than in Canada and other parts of the world. The analyzed costs in the sensitivity analysis were $2,220 [45] as well as a mean of the two costs ($3,330).

4 Results

4.1 Systematic Literature Review

4.1.1 Selection of studies

An overview of the included and excluded articles in every step of the systematic literature review can be seen in Figure 4.1. In total, 2,060 articles were initially extracted from the digital databases. The articles were sorted in folders in Endnote [33], one for every test. 657 of the extracted studies were in MRI, 533 in PHI, 127 in 4Kscore®, 1 in Prostarix, 4 in ConfirmMDx, 426 in PCA3 and 312 in PCMT. After removing duplicates they were reduced to 2,031 studies in total. In this stage 654 were in MRI, 519 in PHI, 126 in 4Kscore®, 1 in Prostarix, 4 in ConfirmMDx, 415 in PCA3 and 312 in PCMT.

All 2,031 remaining articles were screened for relevant information in title and abstract and 1,905 articles were excluded. The remaining 126 studies were kept for full text assessment. Of these 27 were in MRI, 36 in PHI, 20 in 4Kscore®, 0 in Prostarix, 1 in ConfirmMDx, 42 in PCA3 and 0 in PCMT. 97 articles were excluded while screening the full text articles and thus 29 articles were included for assessment of quality and data extraction. Of these 29, 4 were in MRI, 7 in PHI, 10 in 4Kscore®, 0 in Prostarix, 1 in ConfirmMDx, 7 in PCA3 and 0 in PCMT.

4.1.2 Data extraction and quality assessment

Data was extracted and the quality was assessed using the developed forms in Appendix A and B. During this step, 1 additional article was excluded from PHI and 2 from 4Kscore® due to lack of relevant data. Thus, 26 studies were left for data synthesis as can be seen in the flow chart in Figure 4.1. The assessed articles are presented in Table 4.1.2 with the calculated resulting quality score for each article.

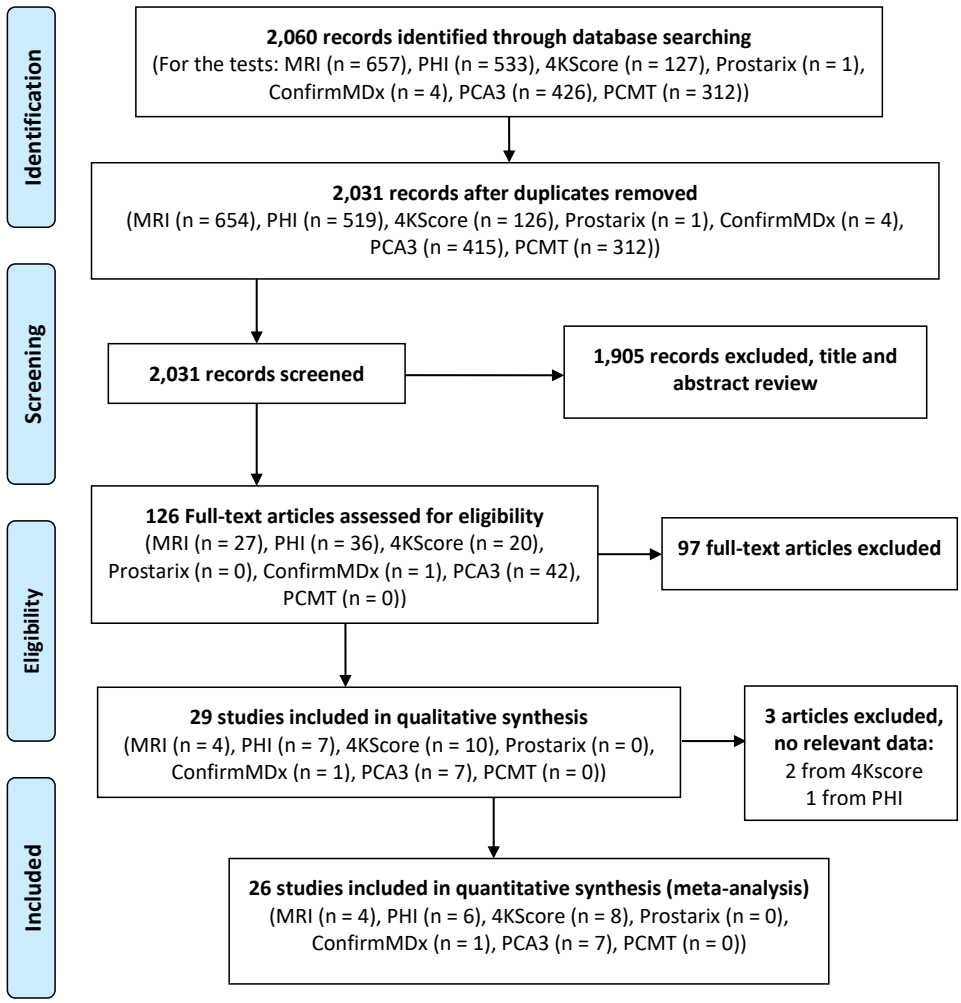


Figure 4.1: Flow diagram over the included and excluded articles in the systematic review. Diagram design based on [46].

Table 4.1: Resulting articles in every test from the systematic literature review with corresponding quality scores

Article title	Reference	Quality (score)
4Kscore®		
A four-kallikrein panel for the prediction of repeat prostate biopsy: data from the European Randomized Study of Prostate Cancer Screening in Rotterdam, Netherlands	[3]	Excellent (0.85)
A panel of kallikrein markers can reduce unnecessary biopsy for prostate cancer: data from the European Randomized Study of Prostate Cancer Screening in Göteborg, Sweden	[36]	Excellent (0.83)
Reducing Unnecessary Biopsy During Prostate Cancer Screening Using a Four-Kallikrein Panel: An Independent Replication	[38]	Excellent (0.82)
Impact of Recent Screening on Predicting the Outcome of Prostate Cancer Biopsy in Men With Elevated Prostate-Specific Antigen	[47]	Excellent (0.80)
The 4Kscore®Test Reduces Prostate Biopsy Rates in Community and Academic Urology Practices	[48]	Good (0.64)
A panel of kallikrein markers can predict outcome of prostate biopsy following clinical work-up: an independent validation study from the European Randomized Study of Prostate Cancer screening, France	[49]	Good (0.63)
A Four-kallikrein Panel Predicts High-grade Cancer on Biopsy: Independent Validation in a Community Cohort	[50]	Good (0.62)

A Four-Kallikrein Panel Predicts Prostate Cancer in Men with Recent Screening: Data from the European Randomized Study of Screening for Prostate Cancer, Rotterdam	[37]	Good (0.56)
MRI		
In-parallel comparative evaluation between multiparametric magnetic resonance imaging, prostate cancer antigen 3 and the prostate health index in predicting pathologically confirmed significant prostate cancer in men eligible for active surveillance	[51]	Excellent (0,89)
Role of Magnetic Resonance Imaging in Prostate Cancer Screening: A Pilot Study Within the Göteborg Randomised Screening Trial	[52]	Excellent (0.79)
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	[53]	Good (0.65)
The value of endorectal MR imaging to predict positive biopsies in clinically intermediate-risk prostate cancer patients	[54]	Poor (0.38)
PHI		
The Prostate Health Index in predicting initial prostate biopsy outcomes in Asian men with prostate-specific antigen levels of 4–10 ng/mL	[55]	Excellent (0.93)
The impact of baseline [-2]proPSA-related indices on the prediction of pathological reclassification at 1 year during active surveillance for low-risk prostate cancer: the Japanese multicenter study cohort	[56]	Excellent (0.77)

Clinical performance of serum prostate specific antigen isoform [-2]proPSA (p2PSA) and its derivatives, %p2PSA and the prostate health index (PHI), in men with a family history of prostate cancer: results from a multicentre European study, the PROMEtheuS project	[41]	Excellent (0.75)
Serum isoform [-2]proPSA derivatives significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2–10 ng/ml: a multicentric European study.	[42]	Good (0.70)
Improving multivariable prostate cancer risk assessment using the Prostate Health Index	[57]	Good (0.67)
Clinical utility of %p2PSA and prostate health index in the detection of prostate cancer	[58]	Good (0.64)
PCA3		
PCA3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study.	[44]	Excellent (0.82)
Impact of adoption of a decision algorithm including PCA3 for repeat biopsy on the costs for prostate cancer diagnosis in France	[59]	Good (0.70)
Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy	[60]	Good (0.68)
Diagnostic performance of PCA3 to detect prostate cancer in men with increased prostate specific antigen: a prospective study of 1,962 cases.	[61]	Good (0.68)
Clinical judgment versus biomarker prostate cancer gene 3: which is best when determining the need for repeat prostate biopsy?	[43]	Good (0.64)

Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions	[62]	Good (0.58)
Biopsy and treatment decisions in the initial management of prostate cancer and the role of PCA3; a systematic analysis of expert opinion.	[63]	Good (0.50)
ConfirmMDx		
Reduced Rate of Repeated Prostate Biopsies Observed in ConfirmMDx Clinical Utility Field Study	[64]	Good (0.60)

4.1.3 Data synthesis

The extracted data was analyzed and values for the base case model were chosen from the articles with the best available data. When no data existed or the articles had low quality, data was used from the TRUSGB case in the previous study [6] if applicable. The chosen values and corresponding references are presented in Table 4.2 and 4.3. The rate of biopsies presented above the test specific rates in Table 4.3 is the common rate of patients referred to a repeat biopsy after an initial negative result. This rate was not extracted from an article in the systematic literature review since is is not specific for any of the tests and was not presented in any of the included studies.

Table 4.2: Rates with main and confirmatory references for the screening tests

Rates	Base Case	Reference	Confirmatory Reference
4Kscore®			
Rate of biopsies after positive 4Kscore®	48.7%	[38]	[49]
Rate of low-risk prostate cancers after positive biopsy	58.3%	[38]	[49]
Rate of significant (intermediate/high-risk) prostate cancers after positive biopsies	41.7%	[38]	[49]
Rate of false negative 4Kscore®	12.9%	[38]	[49]
PHI			
Rate of biopsies after positive PHI	48.0%	[42]	[41]
Rate of false negative PHI	29.2%	[42]	[41]

Table 4.3: Rates with main and confirmatory references for the new tests in the phase after a negative biopsy

Rates	Base Case	Reference	Confirmatory Reference
Rate of repeat biopsies after a 3 year period	43.0%	[65]	
Calculated yearly rate of patients referred for a repeat biopsy	17.1%	[65]	
PCA3			
Rate of repeat biopsies after PCA3	50.4%	[44]	[43]
Rate of positive repeat biopsies after PCA3	33.6%	[44]	
Rate of false negative PCA3	10.0%	[44]	[43], [61]
ConfirmMDx			
Rate of repeat biopsies after ConfirmMDx	4.3%	[64]	
Rate of positive repeat biopsies after ConfirmMDx	0.0%	[64]	
Rate of false negative ConfirmMDx	10.0%	[64]	

Table 4.4: Test costs in Canadian dollars

Test	Cost	Reference	Confirmatory Reference
4Kscore®	$800	[66]	[45]
PHI	$150	[39]	[40]
PCA3	$385	[67]	[59]
ConfirmMDx	$4,440	[40]	[45]

4.2 Modeling

There were two resulting model structures, one for the screening tests and one for the tests after a negative biopsy. The state transition diagrams for the two TreeAge models can be seen in Figure 4.2 and 4.3. The diagrams are a visual representation of the models with all available transitions, from screening until end-of-life. The modeled tests were PHI and 4Kscore® for the screening phase and PCA3 and ConfirmMDx for the phase after a negative initial biopsy.

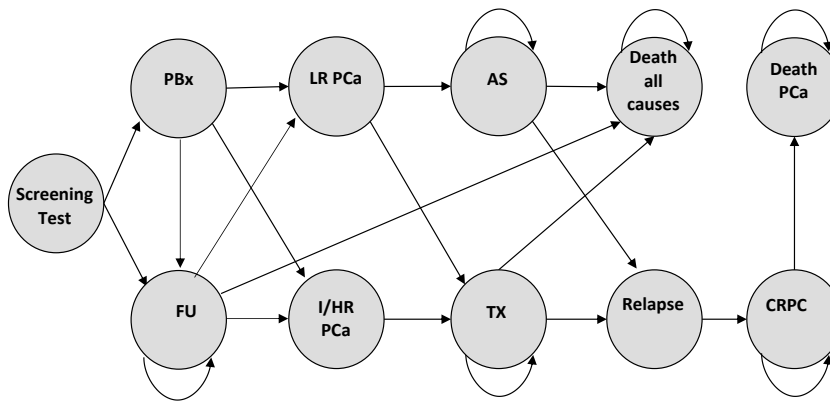


Figure 4.2: The resulting screening TreeAge model visualized as a state transition diagram. PBx = Prostate Biopsy, FU = Follow Up, LR PCa = Low Risk Prostate Cancer, I/HR PCa = Intermediate/High Risk Prostate Cancer, AS = Active Surveillance, TX = Treatment and CRPC = Castration-Resistant Prostate Cancer.

The main difference in the screening model compared to the previously created model (Appendix E) is the initial stratification of patients. Instead of referring all patients to an initial biopsy as in the previous model, only the patients with an increased risk for prostate cancer according to the screening test are biopsied. The nodes for this stratification in the TreeAge model can be seen in Figure 4.4 and 4.5. In the state transition diagram in Figure 4.2, the stratification is from the state Screening Test to the states PBx and FU.

The main change in the model for introducing a test after a negative biopsy was an added branch for a repeat biopsy, see Figure 4.6 and 4.7. This branch harbors the patients that are referred for a second biopsy after an initial negative. This is where the new test is performed and the patients are re-stratified to perform a second biopsy or not depending on the test result. In the state transition diagram in Figure 4.3, these events are located in the stratification from PBx - to Repeat PBx or FU. The examples in Figure 4.4 -

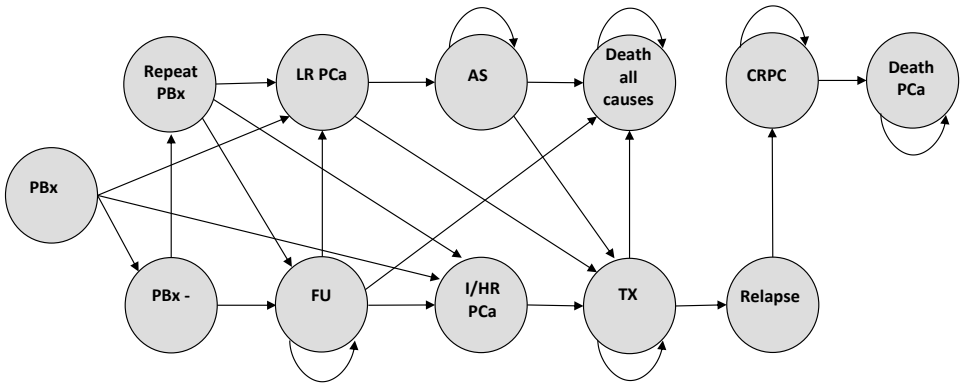


Figure 4.3: The resulting TreeAge model after a negative initial biopsy visualized as a state transition diagram. PBx = Prostate Biopsy, PBx- = Negative biopsy result, FU = Follow Up, LR PCa = Low Risk Prostate Cancer, I/HR PCa = Intermediate/High Risk Prostate Cancer, AS = Active Surveillance, TX = Treatment and CRPC = Castration-Resistant Prostate Cancer.

4.7 do not show the full models, only main changes for every assessed test compared to the initial model.

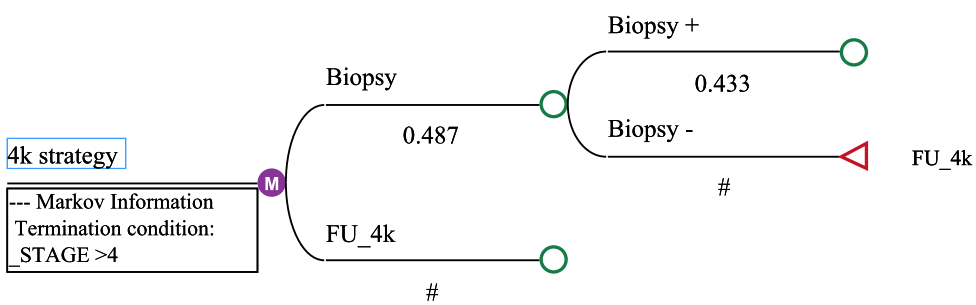


Figure 4.4: Additional nodes in the 4Kscore® TreeAge Model and corresponding transition probabilities.

An overview of the resulting costs and QALY for the evaluated tests in both phases as well as the corresponding results from the previous study can be seen in Table 4.5. The costs and QALY are incremental, calculated for 10, 15 and 20 years respectively. The standard errors (standard deviation of the mean) of the values are presented along with the results in parentheses.

Table 4.5 also presents dominant values (lower costs and higher QALY) for the evaluated tests compared to the strategies in the previous study. Dominant values compared to the corresponding values for TRUSGB or MRGB are marked with a t or an m, respectively. Results for 5 years are presented in

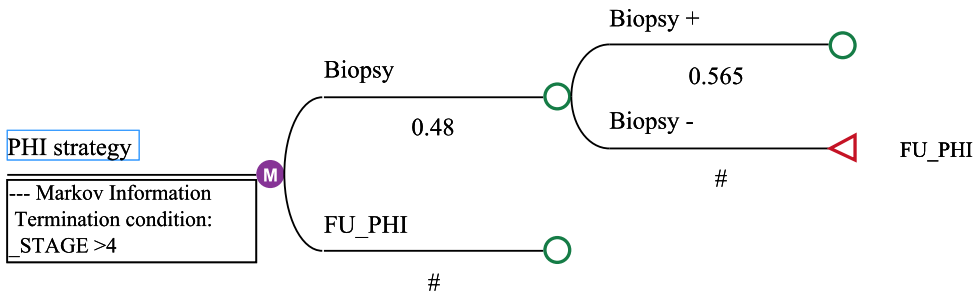


Figure 4.5: Additional nodes in the PHI TreeAge Model and corresponding transition probabilities.

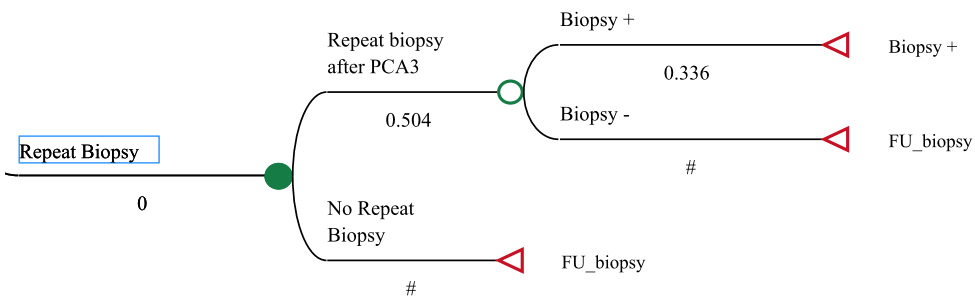


Figure 4.6: Additional nodes in the PCA3 TreeAge Model and corresponding transition probabilities.

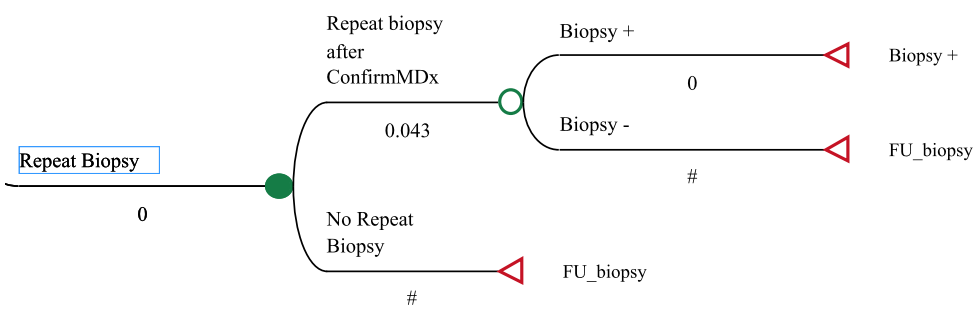


Figure 4.7: Additional nodes in the ConfirmMDx TreeAge Model and corresponding transition probabilities.

Appendix C for the new tests and in Appendix H for TRUSGB and MRGB. These are not presented in this section since 5 years is a very short time horizon in prostate cancer and the results are similar to the results for 10, 15 and 20 years.

Table 4.5: Incremental cost and QALY summary table for 10, 15 and 20 years. Standard error (SE) in parentheses. Values dominating the TRUSGB strategy marked with a t, values dominating the MRGB strategy marked with an m.

Test	Cost 10 years (SE)	QALY 10 years (SE)	Cost 15 years (SE)	QALY 15 years (SE)	Cost 20 years (SE)	QALY 20 years (SE)
Screening tests						
4Kscore®	$\$8{,}300^{tm}$ (99)	7.42^{tm} (0.01)	$\$10{,}389^{tm}$ (146)	9.42^{tm} (0.02)	$\$11{,}995^{tm}$ (178)	10.59^{tm} (0.03)
PHI	$\$11{,}164^{t}$ (115)	7.28^{t} (0.01)	\$14,700 (171)	9.18^{t} (0.02)	\$17,419 (210)	10.25^{t} (0.02)
Tests after negative biopsy						
PCA3	\$11,525 (117)	7.24 (0.01)	\$14,951 (184)	9.12 (0.02)	\$17,480 (221)	10.21 (0.03)
ConfirmMDx	\$11,706 (106)	7.24 (0.01)	\$15,092 (147)	9.13 (0.02)	\$17,598 (179)	10.21 (0.03)
Data from previous study						
TRUSGB	\$11,526 (103)	7.22 (0.01)	\$14,954 (143)	9.11 (0.02)	\$17,495 (169)	10.19 (0.03)
MRGB	\$10,011 (128)	7.31 (0.01)	\$12,814 (185)	9.25 (0.02)	\$14,866 (219)	10.37 (0.03)

Overall, the modeled screening tests showed potential of a better cost-effectiveness compared to the tests after a negative biopsy for 5, 10, 15 and 20 years. The dominating strategies between the tests and TRUSGB are summarized in Table 4.6, while the tests are compared to MRGB in Table 4.7. The strategies are considered equal if neither the costs nor effectivenesses are significantly different.

Compared to TRUSGB, the screening tests were dominating while the tests after a negative biopsy were considered equal or dominated. However, the differences were very small between the TRUSGB strategy and PHI, PCA3 as well as ConfirmMDx, both in cost and in QALY. PHI was only

dominating in both cost and QALY for a 10 year period, while there was no significant difference in cost for 15 and 20 years.

Among the new tests researched in this project, 4Kscore® showed the best potential of increasing cost-effectiveness. The results for the 4Kscore® are dominating when compared to the results for TRUSGB and MRGB (Appendix G) from the previous study. 4Kscore® is the only researched test with a better cost-effectiveness compared to MRGB. Analyzing the results from the tests, 4Kscore® has potential for the best cost-effectiveness progression over a long time horizon.

Table 4.6: Dominant cost-effectiveness strategy table, new tests compared to TRUSGB strategy

Test	10 years	15 years	20 years
Screening tests			
4Kscore®	**Dominating**	**Dominating**	**Dominating**
PHI	**Dominating**	**Dominating**	**Dominating**
Tests after negative biopsy			
PCA3	Equal	Equal	Equal
ConfirmMDx	Dominated	Equal	Equal

Table 4.7: Dominant cost-effectiveness strategy table, new tests compared to MRGB strategy

Test	10 years	15 years	20 years
Screening tests			
4Kscore®	**Dominating**	**Dominating**	**Dominating**
PHI	Dominated	Dominated	Dominated
Tests after negative biopsy			
PCA3	Dominated	Dominated	Dominated
ConfirmMDx	Dominated	Dominated	Dominated

4.3 Sensitivity Analysis

The resulting incremental costs and QALY in the sensitivity analysis of the TreeAge models are presented in Table 4.8 for the 4Kscore® test, 4.9 for PHI, 4.10 for PCA3 and 4.11 for ConfirmMDx. The tables include sensitivity analyses for 10, 15 and 20 years while the corresponding analyses for 5 years are presented in Appendix C. The sensitivity analyses for TRUSGB and MRGB from the previous study are presented in Appendix G and H.

Every resulting value in the sensitivity analyses was compared to the corresponding values for TRUSGB and MRGB in Appendix G. If a value for a test was dominant compared to the corresponding value in the TRUSGB sensitivity analysis it is marked with a t in the table and if it was dominant compared to the corresponding MRGB value it is marked with an m. Some sensitivity analyses were not performed for all tests and thus these values have not been compared. The statistical significance of the differences were not researched and these results should only be regarded as indications.

In the sensitivity analysis, 4Kscore® was still the dominant strategy comparing to both previously researched methods for all values where comparisons were possible.

Table 4.8: Sensitivity Analysis table using 4Kscore® as screening test, costs and utilities are incremental. If the value is dominant to the corresponding value for TRUSGB or MRGB it is marked with a t or an m, respectively

Variable and variations	Cost (10 years)	QALY (10 years)	Cost (15 years)	QALY (15 years)	Cost (20 years)	QALY (20 years)
Variation of discount rate						
No discount rate	$\$9{,}790^{tm}$	9.06^{tm}	$\$13{,}515^{tm}$	12.61^{tm}	$\$17{,}205^{tm}$	15.27^{tm}
3% discount rate	$\$8{,}827^{tm}$	8.01^{tm}	$\$11{,}437^{tm}$	10.51^{tm}	$\$14{,}030^{tm}$	12.35^{tm}
10% discount rate	$\$7{,}180^{tm}$	6.24^{tm}	$\$8{,}372^{tm}$	7.39^{tm}	$\$9{,}114^{tm}$	7.93^{tm}
Variation of AS rate						
10%	$\$8{,}376^{tm}$	7.41^{tm}	$\$10{,}487^{tm}$	9.40^{tm}	$\$12{,}093^{tm}$	10.57^{tm}
20%	$\$8{,}181^{tm}$	7.42^{tm}	$\$10{,}218^{tm}$	9.42^{tm}	$\$11{,}787^{tm}$	10.59^{tm}
25%	$\$8{,}090^{tm}$	7.43^{tm}	$\$10{,}145^{tm}$	9.43^{tm}	$\$11{,}718^{tm}$	10.60^{tm}
40%	$\$7{,}793^{m}$	7.45^{m}	$\$9{,}798^{m}$	9.45^{m}	$\$11{,}328^{m}$	10.63^{m}
50%	$\$7{,}598^{m}$	7.46^{m}	$\$9{,}571^{m}$	9.47^{m}	$\$11{,}068^{m}$	10.65^{m}
Variation of probability of recurrence in intermediate-high-risk group						
2.9%	$\$7{,}939^{tm}$	7.43^{tm}	$\$9{,}728^{tm}$	9.44^{tm}	$\$11{,}119^{tm}$	10.62^{tm}
5.3%	$\$8{,}459^{tm}$	7.41^{tm}	$\$10{,}661^{tm}$	9.39^{tm}	$\$12{,}333^{tm}$	10.55^{tm}
6.5%	$\$8{,}710^{tm}$	7.41^{tm}	$\$11{,}122^{tm}$	9.38^{tm}	$\$12{,}932^{tm}$	10.54^{tm}

7.7%	$8,953^{tm}	7.40^{tm}	$11,557^{tm}	9.37^{tm}	$13,465^{tm}	10.51^{tm}
Variation of utilities						
0.95; 0.95; 0.81; 0.48	$8,282^{tm}	7.49^{tm}	$10,355^{tm}	9.51^{tm}	$11,975^{tm}	10.70^{tm}
0.89; 0.89; 0.75; 0.42	$8,282^{tm}	7.34^{tm}	$10,355^{tm}	9.31^{tm}	$11,975^{tm}	10.46^{tm}
0.85; 0.85; 0.70; 0.40	$8,282^{tm}	7.24^{tm}	$10,355^{tm}	9.17^{tm}	$11,975^{tm}	10.46^{tm}
1.00; 1.00; 1.00; 1.00	$8,282^{tm}	7.66^{tm}	$10,355^{tm}	9.78^{tm}	$11,975^{tm}	11.03^{tm}
Variation of Follow Up cost						
$200.0	$7,939	7.42	$10,012	9.41	$11,632	10.58
$232.5	$8,067	7.42	$10,140	9.41	$11,760	10.58
$395.0	$8,708	7.42	$10,781	9.41	$12,401	10.58
Variation of initial positive biopsy probability						
33.0%	$7,833	7.44	$9,776	9.45	$11,276	10.63
56.5%	$8,864	7.38	$11,103	9.35	$12,831	10.50
65.0%	$9,226	7.37	$11,616	9.33	$13,442	10.48
Variation of rate of initial biopsies						
40.0%	$7,886	7.44	$9,859	9.44	$11,368	10.62
63.8%	$8,998	7.38	$11,272	9.35	$13,008	10.50

Table 4.9: Sensitivity Analysis table using PHI as screening test, costs and utilities are incremental. If the value is dominant to the corresponding value for TRUSGB or MRGB it is marked with a t or an m, respectively

Variable and variations	Cost (10 years)	QALY (10 years)	Cost (15 years)	QALY (15 years)	Cost (20 years)	QALY (20 years)
Variation of discount rate						
No discount rate	$\$13{,}417^t$	8.87^t	\$19,782	12.24^t	\$25,954	14.67^t
3% discount rate	$\$11{,}983^t$	7.85^t	$\$16{,}444^t$	10.22^t	\$20,190	11.70^t
10% discount rate	$\$9{,}527^t$	6.13^t	$\$11{,}564^t$	7.22^t	$\$12{,}804^t$	7.72^t
Variation of AS rate						
10%	$\$11{,}264^t$	7.27^t	$\$14{,}817^t$	9.16^t	\$17,940	10.37^{tm}
20%	$\$11{,}080^t$	7.28^t	$\$14{,}623^t$	9.17^t	$\$17{,}352^t$	10.25^t
25%	$\$10{,}986^t$	7.29^t	$\$14{,}506^t$	9.18^t	$\$17{,}174^t$	10.26^t
40%	\$10,733	7.31	\$14,226	9.21	\$16,894	10.29
50%	\$10,534	7.32^m	\$13,996	9.23	\$16,616	10.32
Variation of probability of recurrence in intermediate-high-risk group						
2.9%	$\$10{,}367^t$	7.30^t	$\$13{,}182^t$	9.22^t	$\$15{,}362^t$	10.34^t
5.3%	$\$11{,}545^t$	7.26^t	$\$15{,}397^t$	9.13^t	$\$18{,}311^t$	10.19^t

6.5%	\$12,113^{t}	7.25^{t}	\$16,436^{t}	9.10^{t}	\$19,611^{t}	10.14^{t}
7.7%	\$12,631^{t}	7.23^{t}	\$17,295^{t}	9.06^{t}	\$20,681^{t}	10.07^{t}
Variation of utilities						
0.95; 0.95; 0.81; 0.48	\$11,171^{t}	7.39^{t}	\$14,715^{t}	9.33^{t}	\$17,424^{t}	10.43^{t}
0.89; 0.89; 0.75; 0.42	\$11,171^{t}	7.16^{t}	\$14,715^{t}	9.00^{t}	\$17,424^{t}	10.05^{t}
0.85; 0.85; 0.70; 0.40	\$11,171^{t}	6.99^{t}	\$14,715^{t}	8.78^{t}	\$17,424^{t}	9.79^{t}
1.00; 1.00; 1.00; 1.00	\$11,171^{t}	7.67^{tm}	\$14,715^{t}	9.76^{tm}	\$17,424^{t}	10.97^{tm}
Variation of Follow Up cost						
\$200.0	\$10,954	7.27	\$14,498	9.16	\$17,207	10,24
\$232.5	\$11,035	7.27	\$14,579	9.16	\$17,288	10,24
\$395.0	\$11,440	7.27	\$14,983	9.16	\$17,692	10,24
Variation of test cost						
\$410	\$11,431	7.27	\$14,975	9.16	\$17,684	10,24
\$670	\$11,691	7.27	\$15,235	9.16	\$17,944	10,24
Variation of rate of initial biopsies						
66.25%	\$11,936	7.23	\$15,235	9.11	\$18,471	10.17
84.50%	\$12,765	7.19	\$16,756	9.04	\$19,712	10.08

Variation of false negative rate						
10.4%	$8,197	7.38	$10,478	9.36	$12,211	10.51
19.8%	$9,824	7.33	$12,815	9.25	$15,073	10.36
Variation of cut-off, FN = false negative						
$27.6 (performed biopsies 84.5% and FN 26%)	$12,478	7.20	$16,342	9.05	$19,240	10.10
$40.3 (performed biopsies 44.3% and FN 28.4%)	$10,948	7.29	$14,395	9.18	$17,053	10.26
$50.9 (performed biopsies 25.5% and FN 33.1%)	$10,753	7.31	$14,235	9.20	$16,939	10.28

Table 4.10: Sensitivity Analysis table using the PCA3 score after a negative initial biopsy, costs and utilities are incremental. If the value was dominant to the corresponding value for TRUSGB or MRGB it was marked with a t or an m, respectively

Variable and variations	Cost (10 years)	QALY (10 years)	Cost (15 years)	QALY (15 years)	Cost (20 years)	QALY (20 years)
Variation of discount rate						
No discount rate	\$13,645	8.82^{t}	\$21,088	12.75^{tm}	\$25,629	14.62^{t}
3% discount rate	\$12,307	7.80^{t}	\$16,655	10.17^{t}	\$20,167	11.65^{t}
10% discount rate	\$10,039	6.09^{t}	\$12,027	7.18^{t}	\$13,191	7.68^{t}
Variation of AS rate						
10%	\$11,639	7.22^{t}	\$15,059	9.11^{t}	$\$17{,}577^{t}$	10.18^{t}
20%	\$11,471	7.24^{t}	\$14,890	9.12^{t}	\$17,427	10.20^{t}
25%	\$11,395	7.24^{t}	\$14,803	9.13^{t}	\$17,286	10.21^{t}
40%	\$11,107	7.26	\$14,450	9.16	\$16,938	10.25
50%	\$10,935	7.27	\$14,319	9.17	\$16,798	10.26
Variation of probability of recurrence in intermediate-high-risk group						
2.9%	\$10,632	7.26^{t}	$\$13{,}306^{t}$	9.18^{t}	$\$15{,}365^{t}$	10.29^{t}

5.3%	\$11,977	7.22^t	\$15,713	9.09^t	\$18,443	10.14^t
6.5%	\$12,611	7.20^t	\$16,747	9.04^t	\$19,694	10.09^t
7.7%	\$13,208	7.18^t	\$17,744	9.00^t	\$20,953	10.02^t
Variation of utilities						
0.95; 0.95; 0.81; 0.48	\$11,554	7.35^t	\$15,009	9.28^t	\$17,550	10.38^t
0.89; 0.89; 0.75; 0.42	\$11,554	7.11^t	\$15,009	8.95^t	\$17,550	10.01^t
0.85; 0.85; 0.70; 0.40	\$11,554	6.94^t	\$15,009	8.73^t	\$17,550	9.75^t
1.00; 1.00; 1.00; 1.00	\$11,554	7.65^{tm}	\$15,009	9.72^t	\$17,550	10.92^t
Variation of FU cost						
\$200.0	\$11,359	7.23	\$14,814	9.12	\$17,354	10.19
\$232.5	\$11,432	7.23	\$14,887	9.12	\$17,427	10.19
\$395.0	\$11,796	7.23	\$15,251	9.12	\$17,792	10.19
Variation of repeat biopsy rate						
37%	\$11,493	7.23	\$14,892	9.12	\$17,412	10.20
60%	\$11,543	7.23	\$14,949	9.12	\$17,471	10.19

Table 4.11: Sensitivity analysis table using the ConfirmMDx after a negative initial biopsy, costs and QALY are incremental. If the value was dominant to the corresponding value for TRUSGB or MRGB it was marked with a t or an m, respectively

Variable and variations	Cost (10 years)	QALY (10 years)	Cost (15 years)	QALY (15 years)	Cost (20 years)	QALY (20 years)
Variation of discount rate						
No discount rate	\$13,734	8.82^t	\$19,741^t	12.18^t	\$25,410^t	14.63^t
3% discount rate	\$12,420	7.80^t	\$16,631	10.17^t	\$20,072	11.66^t
10% discount rate	\$10,184	6.09^t	\$12,108	7.18^t	\$13,248	7.68^t
Variation of AS rate						
10%	\$11,775	7.22^t	\$15,131	9.11^t	\$17,656	10.19^t
20%	\$11,594	7.24^t	\$14,972	9.13^t	\$17,476	10.20^t
25%	\$11,518	7.24^t	\$14,897	9.13^t	\$17,405	10.22^t
40%	\$11,246	7.26	\$14,581	9.16	\$17,027	10.25
50%	\$11,092	7.28	\$14,378	9.18	\$16,785	10.27
Variation of probability of recurrence in intermediate-high-risk group						
2.9%	\$10,821	7.26^t	\$13,494	9.18^t	\$15,537	10.29^t

5.3%	\$12,082	7.22^t	\$15,763	9.09^t	\$18,392	10.15^t
6.5%	\$12,747	7.20^t	\$16,830	9.05^t	\$19,740	10.09^t
7.7%	\$13,346	7.18^t	\$17,773	9.01^t	\$20,894	10.03^t
Variation of utilities						
0.95; 0.95; 0.81; 0.48	\$11,678	7.35^t	\$15,025	9.28^t	\$17,513	10.38^t
0.89; 0.89; 0.75; 0.42	\$11,678	7.11^t	\$15,025	8.95^t	\$17,513	10.01^t
0.85; 0.85; 0.70; 0.40	\$11,678	6.94^t	\$15,025	8.73^t	\$17,513	9.76^t
1.00; 1.00; 1.00; 1.00	\$11,678	7.64^m	\$15,025	9.71	\$17,513	10.91^t
Variation of FU cost						
200.0	\$11,480	7.23	\$14,826	9.12	\$17,314	10.20
232.5	\$11,554	7.23	\$14,900	9.12	\$17,389	10.20
395.0	\$11,924	7.23	\$15,271	9.12	\$17,759	10.20
Variation of cost of ConfirmMDx test						
\$2,220	\$11,524	7.23	\$14,870	9.12	\$17,358	10.20
\$3,330	\$11,601	7.23	\$14,947	9.12	\$17,436	10.20

5 Discussion

5.1 Systematic Literature Review

5.1.1 Selection of studies

The number and quality of the articles varied between the tests, depending on how long they had been on the market and the present clinical usage of them. As expected, due to the novelty of the tests, the resulting number of articles after the systematic literature review was low. More clinical utility studies of the tests are needed before any definitive conclusions about their usefulness can be drawn. The low amount of articles in all tests limited the data extracted for the models. This simplified the work but it also limited the reliability and robustness of the results.

One difficulty that extended the time spent on the systematic literature review was that clinical utility was mentioned in many articles but with different definitions. Clinical utility is an increasingly popular term [28] even in articles only assessing the performance of tests. A high proportion of the abstracts that were included in the first stage of the systematic literature review were later proven to be irrelevant, often due to only assessing the performance of the tests.

Two of the tests with a low number of included articles were ConfimMDx and PCMT. One reason for the low number of utility studies for these particular tests, except the novelty of the tests, is that they are tissue based. Due to this, data or samples from the initial biopsies are needed to complete the studies. This is done less frequently than blood samples which leads to fewer performed studies.

There is a risk that relevant articles were excluded during the process, particularly during the initial screening due to the high number of articles that were assessed. This risk was minimized by letting two students, one of the authors and one from the team, perform every step individually and discuss all articles based on the inclusion and exclusion criteria.

5.1.2 Data extraction and quality assessment method

Every question in the quality assessment form was graded from 0 to 2 or set to N/A. All questions had the same impact on the final quality score which implies that every question was equally important regarding the quality of the study and the results. This might not have been the case but simplified the process, it was difficult to evaluate the importance of every question since the impact of the different aspects on the result was unknown.

Even if the quality score was not linear, it applied on all studies. For example, this limitation had an impact on the 4Kscore®, where the score was excellent or good but all the articles had biases. This should have affected the scoring more than it did.

The quality assessment was a subjective evaluation based on assumptions for every article. If the article was well written it might have gotten a higher quality because of personal preferences, which introduces bias. Since the articles were scored individually and the result was discussed for each article this bias was reduced, but still present.

When the final quality score was calculated, the questions set to N/A were removed from the scoring. This resulted in every question having a larger impact on the final score compared to if all questions were included.

5.1.3 Limitations from included articles

All data concerning the clinical utility of the new tests was collected from articles included in the systematic literature review. During the review, articles with major shortcomings were excluded. But even if the quality of the included articles was considered as high, the limitations from the studies were inherited as uncertainties in the models and the results in this project.

Screening tests

In the screening phase, the use of two tests was modeled: 4Kscore® and PHI.

All 4Kscore®articles in the screening phase included in the data extraction were from the ERSPC trial [68], which makes the results less reliable. All samples in the trial were frozen and kallikreins, which is one of the biomarkers included in the 4Kscore®, decomposes when stored this way [36]. This is an error that is present in all studies based on the samples.

Another reason that the articles included for data extraction in 4Kscore® is that Hans Lilja, who is a co-writer in all articles, has invested in the test. He has interests in the studies and was the one performing the tests in them which may bias the results.

The resulting articles for PHI presented a range of different cut-offs. Additionally, the studies were assessing the performance of the tests as well as the clinical utility. Although the articles presented relevant numbers for this project, the sources were not optimal and the different cut-offs made it difficult to model.

Tests after negative biopsy

In the phase after a negative biopsy, the use of two tests was modeled: PCA3 and ConfirmMDx.

The rate of repeat biopsies after an initial negative biopsy was extracted from [65]. This data was used as a general rate in both developed models in this phase. The article was not included in the systematic literature review since it was not specific for a test and did not include variables sought for initially. Therefore, it has no confirmatory references and has not been evaluated at the same level as the rest of the data. Limitations from [65] include that number of cores taken in every prostate biopsy was not recorded which might have affected the probability of a repeat biopsy. In addition, prostate biopsies were only recorded in a one year frame after a positive screen such as an elevated PSA which might have led to that some negative initial biopsies were missed.

In the ConfirmMDx model, all data was extracted from [64]. In this study, data was collected from health centers that had ordered ConfirmMDx tests to stratify patients with a previous negative biopsy. However, there is no information on the rate of patients with a negative initial biopsy that took the test. Also, the rate of patients that had a repeat biopsy after the tests were stated, but no information on the reason for it. It might not only have been the test result that changed the decision to perform a second biopsy. Another stated limitation in the study is that the number of included patients is relatively small, and thus the results are only indicative of the potential of the test.

For PCA3, the main data was extracted from [44], where all patients with previous negative biopsies were included. Only the patients with the initial biopsy being negative were modeled in this project, so this might result in a bias in the results but still indicates the potential of the test.

5.2 Modeling

Models were developed for PHI, 4Kscore®, PCA3 and ConfirmMDx. The tests that were excluded after the systematic literature review in the screen-

ing phase were MRI and Prostarix, while PCMT, 4Kscore® and PHI were excluded in the phase after a negative biopsy. Prostarix and PCMT were excluded due to the lack of clinical utility studies after the systematic review for both tests. The other tests had resulting articles but there was not enough data to extract into the models.

5.2.1 Assumptions and limitations in models

The models that were developed in this project are simplifications of how patients transition through health states when there is a suspicion of prostate cancer until end-of-life. When modeling, assumptions need to be made to reduce complexity. The models should be as close to the real scenario as possible, with only assumptions that give an insignificant change in results, but it will still introduce some uncertainty.

During the systematic literature review, data from clinical trials of the new tests was sought for use in the models. High quality data from relevant studies made the results reliable and similar to clinical practice. However, all desired data was not found for the new tests. The tests are new to the market which limits the available data and some studies were performed for purposes that were not relevant to this project. In cases where data was missing, it was used from the previous TRUSGB model. The inherited values that were considered uncertain but relevant for the new tests were researched in the sensitivity analysis. One example is the probability for patients with low risk prostate cancer to get referred to AS after performing one of the new tests. This value is expected to increase after implementations of the new tests due to more reliable diagnoses. Another example is that the distribution of the missed cancers between intermediate/high and low risk are expected to change after implementation of the new tests since less of the intermediate/high risk cancers will be missed.

One assumption in the FU state is that patients are merged when transitioning to this state, from two other states, but considered as a homogeneous group. In the screening test model, FU includes patients that did not get an initial biopsy and patients that had a negative biopsy. In the model after a negative biopsy, it includes both patients with an initial and a second negative biopsy. This affects the probability to get diagnosed with prostate cancer in case it was missed on the initial biopsy. The values that overestimated the costs in this case where chosen in the base case model and in relevant cases it was assessed in the sensitivity analysis.

The test costs were difficult to find during the systematic review due to the novelty of the tests. Some costs were from Canada while others were from the United States or Europe. In cases where multiple costs were collected

for the same test, it could be seen that the costs were similar in Canada and Europe while they were higher in the United States. This was researched in the sensitivity analyses for the applicable tests.

Screening tests

The biggest impact of the assumption in the FU state mentioned earlier was in the PHI results since the test's false negative rate is significantly higher than the false negative rate of a biopsy. The results will due to the assumption be an overestimation in cost and underestimation in QALY.

As mentioned before, a range of cut-offs were presented in the articles which made the data more uncertain. FDA has approved PHI, with recommendation not to biopsy patients with a result under 27 but definitely biopsy with a cut-off of 55. This is a wide range and therefore several cut-offs were tested in the sensitivity analysis. The cut-off 41.5 was used in the base model while other cut-offs were run in the sensitivity analysis to compare the change in cost-effectiveness. No specific cut-off has been approved for 4Kscore®, the most widely used cut-off in the clinical utility studies was used in the model.

Tests after negative biopsy

In the TRUSGB model from the previous study, all repeat biopsies were assumed to be performed in FU. In this project, the first repeat biopsies are instead modeled separately. A better stratification of patients due to the new test and a second biopsy before going into FU will probably result in a lower FU cost per year due to fewer repeat biopsies and other examinations. This was included in the sensitivity analysis for all tests, since the effect will probably be present also in the implementation of screening tests.

The main limitation in the ConfirmMDx model is that there was only one resulting article after the systematic literature review. This article included enough data to be modeled but the data needs to be confirmed for the results to be more reliable. As an example, the article presented that the rate of patients with a positive second biopsy was 0%, which is unlikely.

5.3 Sensitivity analysis

The sensitivity analysis was run fewer separate times than the base case, 10 times compared to 100 times, due to time constrains. The confidence intervals were not considered in the comparisons because of the same reason. Therefore, the comparisons of the costs and QALY in the sensitivity analysis are only indications.

The values that were altered in the sensitivity analysis in the previous study were included to be able to compare the results. Two AS rates (40% and 50%) were added in the sensitivity analyses since the stratification in low and intermediate/high risk prostate cancer most likely will be more reliable. Due to this, a higher proportion of the low risk cancers will not be treated with any radical treatment until they show signs of progression.

Utility 1 for all states were added in the sensitivity analysis to compare to if the QALY would be considered to be the same regardless of the state. This scenario was modeled to compare the outcome from the models without the uncertainty in QALY definitions.

The performed sensitivity analyses in this project were deterministic, only altering one variable at a time. It was done using the same method as in the previous study for comparability. The values in the sensitivity analysis were extracted from the literature when applicable. If there was no available data the values were increased or decreased in reasonable ranges. These *reasonable ranges* were set by referring to values for other tests or from the previous study.

Several variables in the models depend on each other and analyzing them in combination in a probabilistic analysis would probably have shown possible alterations better. Automatized methods such as tornado diagrams would also have made the process easier and more accurate. This would have been done in case of more time for the project.

5.4 Results

The following section includes discussion regarding the results for the tests. The results were compared with the results on TRUSGB and MRGB from the previous study. Differences for the base cases and the sensitivity analyses for each test were evaluated.

5.4.1 4Kscore®

The cost-effectiveness of the 4Kscore® dominates both TRUSGB and MRGB for all time horizons. This applies for both the base case and all variables in the sensitivity analyses. One thing that can be pointed out when comparing the differences between the time horizons is that 4Kscore® has a better progress over time than TRUSGB and MRGB.

Further, all specific values evaluated for the test in the sensitivity analysis showed a higher cost-effectiveness compared to the base cases for TRUSGB

and MRGB. According to the results, the 4Kscore® would be the dominating test even if these variables are too uncertain.

The 4Kscore® test had the most resulting extracted data and thus the results for this test are probably most reliable, although the articles had some biases which could have resulted in overestimated benefits for the test. Furthermore, some biases also exist in the model which might have an impact on the results.

The quality of the article chosen for the base case was considered excellent while the confirmatory was considered good. This confirms that the results are reliable, even though some biases in the articles are significant.

5.4.2 PHI

The results from the PHI base case model showed a cost-effectiveness that slightly dominates the TRUSGB strategy but is dominated by the MRGB strategy. The differences between the time horizons were comparable with TRUSGB which implies that the benefits will not increase by much when using PHI over a longer time period.

Comparing the sensitivity analysis for PHI with the corresponding values for MRGB showed that MRGB was still dominating in cost-effectiveness for every value. Comparing to TRUSGB, PHI was dominating in most values, except for costs for when discount rate was set to zero and 3% as well as the AS probability set to 10% with a time horizon of 20 years.

If the rate of patients with low risk prostate cancer going to AS would increase because of the test, the cost-effectiveness would increase using the tests compared to both TRUSGB and MRGB. This scenario is likely compared to TRUSGB, but MRGB has potential of detecting cancers more accurately as well.

The false negative value for PHI, 29.2%, was significantly higher than the false negative rate for TRUSGB, 10.4%, which as mentioned earlier also affects the number of missed cancers in FU. When using the false negative rate for TRUSGB, the cost-effectiveness for PHI was clearly dominating both TRUSGB and MRGB if compared with their base case. A more reasonable rate that was tested was a mean of these values. This resulted in PHI dominating TRUSGB and being equal to MRGB.

When performing the sensitivity analysis with different cut-offs for PHI, the cost-effectiveness did not change much. This implies that the exact value for this variable is irrelevant in the model.

The main article used for data extraction in PHI had a quality score that was considered as good. The confirmatory reference had a score that was

similar, but considered as excellent. The quality scores and cut-offs for these articles were very similar, but the article with a slightly lower quality score was chosen since it was stated that the cut-off was the best combination of specificity and sensitivity.

5.4.3 PCA3

In the base case results, the cost-effectiveness for PCA3 and TRUSGB were considered equal while it was dominated by MRGB. This was the case for all time horizons that were investigated in this project.

When comparing the sensitivity analyses of the test and TRUSGB, there is a indication that the PCA3 test has the potential to increase the QALY more than what the costs decrease. The recommendations on implementation of the test then depends on what is considered most important: the quality of life for the patients or the costs for the health care system.

The variation in the sensitivity analysis that seems to make the PCA3 test dominate TRUSGB regarding cost-effectiveness is increasing the AS rate to 40 or 50%. Studies covering longer time horizons where the treatment stratifications are recorded are needed to evaluate this further.

Overall, the variations of variables in the sensitivity analysis of PCA3 resulted in very small variations in the resulting cost-effectiveness. This indicates that the values used in the models are certain enough for these comparisons.

The quality of the reference used for data extraction to the PCA3 model was considered as excellent, while the confirmatory references were scored as good. This reinforces the results regarding this test.

5.4.4 ConfirmMDx

The results for ConfirmMDx for the base case show a similar cost-effectiveness as TRUSGB, except for 10 years where TRUSGB is dominating. This indicates that ConfirmMDx has a potential over a long time horizon, which is a relevant case for prostate cancer since it is a slow disease. However, the differences are so small that it is impossible to draw any conclusions. Compared to the MRGB strategy, the test is dominated in all cases. The same conclusion was drawn after comparing the sensitivity analyses.

The results in the sensitivity analysis were also compared to TRUSGB to investigate the differences further. Just as in PCA3, the test has better potential in increasing the QALY than the cost.

One uncertain variable for ConfirmMDx that might impact the results is the test cost, which was extracted from a source from the United States

and is much higher than the costs for the other tests. The resulting cost-effectiveness was still very similar to TRUSGB after analyzing an other lower price found in the literature. The cost-effectiveness of the test was not very dependent on the price of it.

Another scenario that was researched in the sensitivity analysis was that a higher proportion of the patients will get referred to AS after implementation of ConfirmMDx. If the AS rate increases to 40 or 50%, compared to 15% with TRUSGB, the cost-effectiveness for the ConfirmMDx test would dominate TRUSGB for all time horizons. However, 40 and 50% are not values extracted from the literature and clinical utility studies on the subject are needed to confirm these changes.

There was only one resulting article to extract data from in ConfirmMDx. The quality of the article was considered good, but not excellent, and the data that could be used in the model was very limited. All values in the ConfirmMDx model are uncertain but since there were no confirmatory articles, there were no indications on what values could be included in the sensitivity analysis apart from the cost. Therefore, these were left until more clinical utility studies are published on the test. The cost-effectiveness results for this test have big uncertainties and more studies should be conducted before any further conclusions can be made.

5.4.5 Comparisons between new tests

The test that proved the best resulting cost-effectiveness was the 4Kscore®. This is also the only new test that dominated the results of MRGB from the previous study. The test with the least favourable cost-effectiveness was the ConfirmMDx test.

As can be seen in Table 4.6 and 4.7, the screening tests overall showed greater potential in decreasing costs and increasing QALY than the tests after a negative biopsy. The tests were evaluated in two different models, one for each phase, and one explanation for this difference is bias in model design. The screening model was mainly developed by E. Bonnevier while E. Palenius was responsible for modelling the tests after a negative biopsy and this might have impacted. However, the differences in results between PHI, PCA3 and ConfirmMDx are very small and a more fair point of view is that 4Kscore® showed potential of increasing the cost-effectiveness compared to TRUSGB and MRGB while the other tests all have similar cost-effectiveness as each other and TRUSGB.

The quality of the articles that data was extracted from was considered as good or excellent in all cases for the four tests. The only test without any extracted data that was considered excellent was ConfirmMDx. With a

higher number of studies with excellent quality the evidence of the results would be stronger, but the quality of the article was high enough for the results to be indicative of the utility of the tests in clinical practice.

The goal was to select values for modeling that rather overestimated the costs and underestimated the QALY than the opposite. This increases the reliability in the potential winnings of introducing 4Kscore® as a screening test. Even the other researched tests have potential because of the small differences in the results but more clinical utility studies are needed.

Not only the benefit in cost-effectiveness is important when implementing new methods. Even if a test shows potential of higher benefit, there might be other disadvantages. For example, the implementation of MRGB entails a better cost-effectiveness compared to most other methods, but not all hospitals have MRI machines. It might be too costly to use the method, even if it is more cost-effective in the long run. A simple blood- or urine test might be easier and cheaper to introduce to the current practise.

There is also a question whether decreased cost or gained QALY is most important. This has an impact in interpretation of the results. The resulting numbers from modeling in this project is not enough to make a decision of which tests should be implemented.

5.5 Sustainability and ethical aspects

The aim in this project was to assess the potential benefit of introducing new tests to the screening phase of prostate cancer. Finding new and more efficient tests is of interest to both clinicians, patients and the society. By introducing new tests the costs can be decreased for everybody involved while the quality of life for the patients is maintained or increased. An increased cost-effectiveness can contribute to a more sustainable society.

Most people can, however, live a normal life with prostate cancer for many years without any symptoms. The patients usually have insignificant cancer or low risk cancer that does not develop before they die of other causes. Treatments might do more harm than helping these patients. Examinations and treatments might cause severe side effects and the knowledge of having cancer might result in problems with anxiety or depression. On the other hand, it is often too late to treat the cancer if treatment is not started until there are symptoms. A controversial question is how many false diagnoses and suffering patients one saved life is worth.

The results from the models are guidelines to use when incorporating new tests into clinical practice. The results are limited and some factors are not taken into consideration. Ethics can be difficult to consider in models while

it is highly relevant in practice. The balance between costs and quality of life is not easy to determine and many find it hard to believe that a human life can have a value measured in money. However, the resources are not infinite and decisions need to be made.

5.6 Future work

More clinical utility studies for the new tests are needed to verify the extracted data and to verify the robustness of the tests. The cost-effectiveness for the new tests needs to be evaluated further to be able to determine the actual benefit of introducing the new tests to the health care market.

Both the screening tests and tests after a negative initial biopsy are used to stratify patients without a previous prostate cancer diagnosis into risk groups to reduce the number of unnecessary biopsies. This implies that all researched tests could be used in both phases, with the exception being if the tests are based on a previous biopsy tissue sample. It would be interesting to investigate for example PCA3 for screening. But to model other scenarios, more clinical utility studies are needed to present more data that can be assessed.

Implementation of one test does not need to rule out the use of a second one or to use current practice simultaneously. A combination of the tests could be used or different methods might be in use in different health centers. A combination of two test could be more beneficial than just using one and would be interesting to research if more time was available.

Another possible scenario that was not researched in this project is to keep the TRUSGB as the initial biopsy for all referred patients but perform repeat biopsies with MRGB. It might be too costly for some hospitals to perform MRGB on all referred patients on initial biopsy but this solution would have the potential to reduce re-biopsies due to the higher specificity in the method. MRGB has shown potential to reduce costs and increase the utility for the patients in previous studies [6] and it would be interesting to research the combination of methods.

Another thing that would be interesting to study is to implement the new tests in other countries. The cost for the tests differ between countries and the health care systems may have different abilities to implement a new intervention.

Even after proving the utility of a new test, it is difficult and time consuming to incorporate new intervention into clinical practice. Clinicians have their way of doing things and it is costly to change a current system. Therefore it is important to have strong proof of clinical utility before changes can

be made. This can be done with more studies regrading the cost-effectiveness and clinical utility.

The results in the project, the values included in the modeling and the chain of events would be interesting to discuss with clinicians to be able to evaluate if the scenarios represents the reality. However, no time was left during this project for this type of validation of results.

6 Conclusions

After the systematic review, PHI and 4Kscore® of the screening tests and PCA3 of the tests after a negative biopsy resulted in the highest number of high quality clinical utility studies. Only one article was included for ConfirmMDx, but it was not excluded due to good quality and relevant results.

4Kscore® had the most extracted data to use in the model and showed a higher cost-effectiveness than both TRUSGB and MRGB. It has the best potential to reduce unnecessary biopsies among the evaluated tests. Over time the benefits of the 4Kscore® increases more than the other methods.

PHI had almost as many included articles after the systematic review as 4Kscore® but the data was more uncertain as many cut-offs was presented. PHI was dominating in cost-effectiveness over TRUSGB but not for MRGB. The number of missed cancers is probably lower than what is modeled and with a lower value PHI has potential to result in a better cost-effectiveness.

PCA3 also had a high number of resulting articles to extract data from. The cost-effectiveness for PCA3 and TRUSGB were considered equal while the test was dominated by MRGB. The test has better potential in increasing the QALY than the cost.

ConfirmMDx was dominated by both TRUSGB and MRGB. The test showed the lowest cost-effectiveness but there were a lot of uncertainties and more clinical utility studies are needed.

The results in this project indicate that it is more cost-effective to use and develop screening test than tests after a negative biopsy. To be able to draw any final conclusions regarding the tests more clinical utility studies and cost-effectiveness studies for all tests are needed.

Bibliography

[1] D. Bayliss, J. Duff, P. Stricker, and K. Walker, "Decision-making in prostate cancer - choosing active surveillance over other treatment options: A literature review," *Urologic Nursing*, vol. 37, no. 1, pp. 15–22, 2017.

[2] A. J. Wein, L. R. Kavoussi, A. C. Novick, A. W. Partin, and C. A. Peters, *Campbell-Walsh Urology: Expert Consult Premium Edition: Enhanced Online Features and Print, 4-Volume Set, 10e (Campbell's Urology (4 Vols.))*. Philadelphia: Saunders, 2011.

[3] A. Gupta, M. J. Roobol, C. J. Savage, M. Peltola, K. Pettersson, P. T. Scardino, A. J. Vickers, F. H. Schroder, and H. Lilja, "A four-kallikrein panel for the prediction of repeat prostate biopsy: data from the european randomized study of prostate cancer screening in rotterdam, netherlands," *British Journal of Cancer*, vol. 103, no. 5, pp. 708–14, 2010.

[4] T. Y. Perez, M. R. Danzig, R. A. Ghandour, K. K. Badani, M. C. Benson, and J. M. McKiernan, "Impact of the 2012 United States Preventive Services Task Force statement on prostate-specific antigen screening: analysis of urologic and primary care practices," *Urology*, vol. 85, no. 1, pp. 85–9, 2015.

[5] M. Hunink, P. Glasziou, J. Siegel, J. Weeks, J. Pliskin, A. Elstein, and M. Weinstein, *Decision making in health and medicine*. New York: Cambridge University Press, 2001.

[6] Y. Cerantola, A. Dragomir, S. Tanguay, F. Bladou, A. Aprikian, and W. Kassouf, "Cost-effectiveness of multiparametric magnetic resonance imaging and targeted biopsy in diagnosing prostate cancer," *Urologic Oncology: Seminars and Original Investigations*, vol. 34, no. 3, pp. 119.e1–119.e9, 2016.

[7] F. Saad and M. McCormack, *Prostate Cancer*, 3rd ed. Montréal, Canada: Annika Parance Publishing, 2012.

[8] J. Schalken, "Molecular and cellular prostate biology: origin of prostate-specific antigen expression and implications for benign prostatic hyperplasia," *BJU International*, vol. 93, no. s1, pp. 5–9, 2004.

[9] Blausen Medical Communications, Inc., "Prostate cancer," Wikimedia Commons, accessed: 2017-06-12.

[10] ——, "Prostate needle biopsy," Wikimedia Commons, accessed: 2017-06-12.

[11] P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Modern Pathology*, vol. 17, no. 3, pp. 292–306, 2004.

[12] L. Zhuang and M. T. Johnson, "How precisely can prostate cancer be managed?" *International Neurourology Journal*, vol. 20, no. 2, pp. 120–130, 2016.

[13] Canadian Cancer Society's Advisory Committee on Cancer Statistics, "Canadian Cancer Statistics 2016," Oct 2016, Toronto, Ontario, ISSN 0835-2976.

[14] Faculty of Radiation Oncology, The Royal Australian and New Zealand College of Radiologists, "External Beam Radiation Therapy (EBRT)," https://www.targetingcancer.com.au/radiation-therapy/ebrt/, accessed: 2017-06-12.

[15] D. W. Lin, L. F. Newcomb, M. D. Brown, D. D. Sjoberg, Y. Dong, J. D. Brooks, P. R. Carroll, M. Cooperberg, A. Dash, W. J. Ellis, M. Fabrizio, M. E. Gleave, T. M. Morgan, P. S. Nelson, I. M. Thompson, A. A. Wagner, and Y. Zheng, "Evaluating the four kallikrein panel of the 4kscore for prediction of high-grade prostate cancer in men in the canary prostate active surveillance study," *European Urology*, 2016.

[16] A. Sreekumar, L. M. Poisson, T. M. Rajendiran, A. P. Khan, Q. Cao, J. Yu, B. Laxman, R. Mehra, R. J. Lonigro, Y. Li, M. K. Nyati, A. Ahsan, S. Kalyana-Sundaram, B. Han, X. Cao, J. Byun, G. S. Omenn, D. Ghosh, S. Pennathur, D. C. Alexander, A. Berger, J. R. Shuster, J. T. Wei, S. Varambally, C. Beecher, and A. M. Chinnaiyan, "Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression," *Nature*, vol. 457, no. 7231, pp. 910–914, 2009.

[17] S. Punnen, N. Pavan, and D. J. Parekh, "Finding the wolf in sheep's clothing: The 4kscore is a novel blood test that can accurately identify the risk of aggressive prostate cancer," *Reviews in Urology*, vol. 17, no. 1, pp. 3–13, 2015.

[18] D. J. Parekh, S. Punnen, D. D. Sjoberg, S. W. Asroff, J. L. Bailen, J. S. Cochran, R. Concepcion, R. D. David, K. B. Deck, I. Dumbadze, M. Gambla, M. S. Grable, R. J. Henderson, L. Karsh, E. B. Krisch, T. D. Langford, D. W. Lin, S. M. McGee, J. J. Munoz, C. M. Pieczonka, K. Rieger-Christ, D. R. Saltzstein, J. W. Scott, N. D. Shore, P. R. Sieber, T. M. Waldmann, F. N. Wolk, and S. M. Zappala, "A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer," *European Urology*, vol. 68, no. 3, pp. 464–470, 2015.

[19] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, Eds., *Magnetic Resonance Imaging*. Wiley-Blackwell, 2014.

[20] J. J. Tosoian, S. Loeb, J. I. Epstein, B. Turkbey, P. L. Choyke, and E. M. Schaeffer, "Active surveillance of prostate cancer: Use, outcomes, imaging, and diagnostic tools," *American Society of Clinical Oncology Educational Book*, vol. 35, pp. e235–e245, 2016.

[21] W. J. Catalona, A. W. Partin, M. G. Sanda, J. T. Wei, G. G. Klee, C. H. Bangma, K. M. Slawin, L. S. Marks, S. Loeb, D. L. Broyles, S. S. Shin, A. B. Cruz, D. W. Chan, L. J. Sokoll, W. L. Roberts, R. H. van Schaik, and I. A. Mizrahi, "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*, vol. 185, no. 5, pp. 1650–1655, 2011.

[22] S. D. Mikolajczyk, K. M. Marker, L. S. Millar, A. Kumar, M. S. Saedi, J. K. Payne, C. L. Evans, C. L. Gasior, H. J. Linton, P. Carpenter, and H. G. Rittenhouse, "A truncated precursor form of prostate-specific antigen is a more specific serum marker of prostate cancer," *Cancer Research*, no. 18, pp. 6958–6963,, sep 2001.

[23] U. S. Food and Drug Administration, "Summary of Safety and Effectiveness Data, Premarket Approval Application Number P090026," https://www.accessdata.fda.gov/cdrh_docs/pdf9/P090026B.pdf, jun 2012, accessed: 2017-05-02.

[24] ——, "Summary of Safety and Effectiveness Data, Premarket Approval Application Number P100033," https://www.accessdata.fda.gov/cdrh_docs/pdf10/P100033B.pdf, feb 2012, accessed: 2017-05-02.

[25] "Physician brochure for the PROGENSA® PCA3 assay," 10210 Genetic Center Drive, San Diego, CA 92121, 2012.

[26] S. M. Falzarano, M. Ferro, E. Bollito, E. A. Klein, G. Carrieri, and C. Magi-Galluzzi, "Novel biomarkers and genomic tests in prostate cancer: a critical analysis," *Minerva urologica e nefrologica*, vol. 67, no. 3, pp. 211–231, 2015.

[27] L. Legisi, E. DeSa, and N. Qureshi, "Use of the prostate core mitomic test in repeated biopsy decision-making: Real-world assessment of clinical utility in a multicenter patient population." *American Health & Drug Benefits*, vol. 9, no. 9, pp. 497–501, 2016.

[28] A. Smart, "A multi-dimensional model of clinical utility," *International Journal for Quality in Health Care*, vol. 18, no. 5, pp. 377–382, 2006.

[29] TreeAge Software, Inc., "TreeAge Pro (R1, 1.0)," 2017, accessed: 2017-02-30.

[30] U. Siebert, O. Alagoz, A. M. Bayoumi, B. Jahn, D. K. Owens, D. J. Cohen, and K. M. Kuntz, "State-transition modeling: A report of the ISPOR-SMDM modeling good research practices task force-3," *Value in Health*, vol. 15, no. 6, pp. 812–820, 2012.

[31] F. A. Sonnenberg and J. R. Beck, "Markov models in medical decision making," *Medical Decision Making*, vol. 13, no. 4, pp. 322–338, 1993, pMID: 8246705.

[32] R. Jain, M. Grabner, and E. Onukwugha, "Sensitivity analysis in cost-effectiveness studies," *Pharmacoeconomics*, vol. 29, no. 4, pp. 297–314, 2011.

[33] Clarivate Analytics, "EndNote (X8.0.1)," 2016, accessed: 2017-02-15.

[34] S. Sommariva, R. Tarricone, M. Lazzeri, W. Ricciardi, and F. Montorsi, "Prognostic value of the cell cycle progression score in patients with prostate cancer: a systematic review and meta-analysis," *European Urology*, vol. 69, no. 1, pp. 107–15, 2016.

[35] T. S. Rector, B. C. Taylor, and T. J. Wilt, "Chapter 12: Systematic review of prognostic tests," *Journal of General Internal Medicine*, vol. 27, no. 1, pp. 94–101, 2012.

[36] A. J. Vickers, A. M. Cronin, G. Aus, C. G. Pihl, C. Becker, K. Pettersson, P. T. Scardino, J. Hugosson, and H. Lilja, "A panel of kallikrein markers can reduce unnecessary biopsy for prostate cancer: data from the european randomized study of prostate cancer screening in goteborg, sweden," *BMC Medicine*, vol. 6, no. 1, 2008.

[37] A. J. Vickers, A. M. Cronin, M. J. Roobol, C. J. Savage, M. Peltola, K. Pettersson, P. T. Scardino, F. H. Schroder, and H. Lilja, "A four-kallikrein panel predicts prostate cancer in men with recent screening: data from the european randomized study of screening for prostate cancer, rotterdam," *Clinical Cancer Research*, vol. 16, no. 12, pp. 3232–9, 2010.

[38] A. Vickers, A. Cronin, M. Roobol, C. Savage, M. Peltola, K. Pettersson, P. T. Scardino, F. Schroder, and H. Lilja, "Reducing unnecessary biopsy during prostate cancer screening using a four-kallikrein panel: an independent replication," *Journal of Clinical Oncology*, vol. 28, no. 15, pp. 2493–8, 2010.

[39] A. Nicholson, J. Mahon, A. Boland, S. Beale, K. Dwan, N. Fleeman, J. Hockenhull, and Y. Dundar, "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 Technol Assess*, vol. 19, no. 87, pp. i–xxxi, 1–191, 2015.

[40] Prostate Conditions Education Council, "Prostate cancer markers," http://www.prostatemarkers.org/images/resources/test-grid.pdf, 2015, accessed: 2017-06-02.

[41] M. Lazzeri, A. Haese, A. Abrate, A. Taille, J. P. Redorta, T. McNicholas, G. Lughezzani, G. Lista, A. Larcher, V. Bini *et al.*, "Clinical performance of serum prostate-specific antigen isoform [-2] propsa (p2psa) and its derivatives,% p2psa and the prostate health index (phi), in men with a family history of prostate cancer: results from a multicentre european study, the prometheus project," *BJU international*, vol. 112, no. 3, pp. 313–321, 2013.

[42] M. Lazzeri, A. Haese, A. de la Taille, J. Palou Redorta, T. McNicholas, G. Lughezzani, V. Scattoni, V. Bini, M. Freschi, A. Sussman, B. Ghaleb, P. Le Corvoisier, J. Alberola Bou, S. Esquena Fernandez, M. Graefen, and G. Guazzoni, "Serum isoform [-2]propsa derivatives significantly improve prediction of prostate cancer at initial biopsy in a total psa range of 2-10 ng/ml: a multicentric european study," *European Urology*, vol. 63, no. 6, pp. 986–94, 2013.

[43] B. Tombal, G. L. Andriole, A. de la Taille, P. Gontero, A. Haese, M. Remzi, M. Speakman, L. Smets, and H. Stoevelaar, "Clinical judgment versus biomarker prostate cancer gene 3: which is best when determining the need for repeat prostate biopsy?" *Urology*, vol. 81, no. 5, pp. 998–1004, 2013.

[44] M. C. Gittelman, B. Hertzman, J. Bailen, T. Williams, I. Koziol, R. J. Henderson, M. Efros, M. Bidair, and J. F. Ward, "Pca3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study," *Journal of Urology*, vol. 190, no. 1, pp. 64–9, 2013.

[45] L. Murphy, M. Prencipe, W. M. Gallagher, and R. W. Watson, "Commercialized biomarkers: new horizons in prostate cancer diagnostics," *Expert Rev Mol Diagn*, vol. 15, no. 4, pp. 491–503, 2015.

[46] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLOS Medicine*, vol. 6, no. 7, pp. 1–6, 2009.

[47] A. J. Vickers, A. M. Cronin, G. Aus, C. G. Pihl, C. Becker, K. Pettersson, P. T. Scardino, J. Hugosson, and H. Lilja, "Impact of recent screening on predicting the outcome of prostate cancer biopsy in men with elevated prostate-specific antigen: data from the european randomized study of prostate cancer screening in gothenburg, sweden," *Cancer*, vol. 116, no. 11, pp. 2612–20, 2010.

[48] B. Konety, S. M. Zappala, D. J. Parekh, D. Osterhout, J. Schock, R. M. Chudler, G. M. Oldford, K. M. Kernen, and J. Hafron, "The 4kscore test reduces prostate biopsy rates in community and academic urology practices," *Reviews in Urology*, vol. 17, no. 4, pp. 231–40, 2015.

[49] A. Benchikh, C. Savage, A. Cronin, G. Salama, A. Villers, H. Lilja, and A. Vickers, "A panel of kallikrein markers can predict outcome of prostate biopsy following clinical work-up: an independent validation study from the european randomized study of prostate cancer screening, france," *Bmc Cancer*, vol. 10, no. 1, 2010.

[50] K. Braun, D. D. Sjoberg, A. J. Vickers, H. Lilja, and A. S. Bjartell, "A four-kallikrein panel predicts high-grade cancer on biopsy: Independent validation in a community cohort," *European Urology*, vol. 69, no. 3, pp. 505–11, 2016.

[51] F. Porpiglia, F. Cantiello, S. De Luca, M. Manfredi, A. Veltri, F. Russo, A. Sottile, and R. Damiano, "In-parallel comparative evaluation between multiparametric magnetic resonance imaging, prostate cancer antigen 3 and the prostate health index in predicting pathologically confirmed significant prostate cancer in men eligible for active surveillance," *BJU International*, vol. 118, no. 4, pp. 527–34, 2016.

[52] A. Grenabo Bergdahl, U. Wilderang, G. Aus, S. Carlsson, J. E. Damber, M. Franlund, K. Geterud, A. Khatami, A. Socratous, J. Stranne, M. Hellstrom, and J. Hugosson, "Role of magnetic resonance imaging in prostate cancer screening: A pilot study within the goteborg randomised screening trial," *Eur Urol*, vol. 70, no. 4, pp. 566–73, 2016.

[53] M. R. Pokorny, M. de Rooij, E. Duncan, F. H. Schroder, R. Parkinson, J. O. Barentsz, and L. C. Thompson, "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," *European Urology*, vol. 66, no. 1, pp. 22–9, 2014.

[54] J. C. Vilanova, J. Comet, A. Capdevila, J. Barcelo, J. L. Dolz, M. Huguet, C. Barcelo, J. Aldoma, and E. Delgado, "The value of endorectal mr imaging to predict positive biopsies in clinically intermediate-risk prostate cancer patients," *European Radiology*, vol. 11, no. 2, pp. 229–235, 2001.

[55] C. F. Ng, P. K. Chiu, N. Y. Lam, H. C. Lam, K. W. Lee, and S. S. Hou, "The prostate health index in predicting initial prostate biopsy outcomes in asian men with prostate-specific antigen levels of 4-10 ng/ml," *International Urology Nephrology*, vol. 46, no. 4, pp. 711–7, 2014.

[56] H. Hirama, M. Sugimoto, K. Ito, T. Shiraishi, and Y. Kakehi, "The impact of baseline [-2]propsa-related indices on the prediction of pathological reclassification at 1 year during active surveillance for low-risk prostate cancer: the japanese multicenter study cohort," *Journal of Cancer Research Clinical Oncology*, vol. 140, no. 2, pp. 257–63, 2014.

[57] R. W. Foley, L. Gorman, N. Sharifi, K. Murphy, H. Moore, A. V. Tuzova, A. S. Perry, T. B. Murphy, D. J. Lundon, and R. W. Watson, "Improving multivariable prostate cancer risk assessment using the prostate health index," *BJU International*, vol. 117, no. 3, pp. 409–17, 2016.

[58] X. Filella, L. Foj, J. M. Auge, R. Molina, and J. Alcover, "Clinical utility of %p2psa and prostate health index in the detection of prostate cancer," *Clinical Chemistry Laboratory Medicine*, vol. 52, no. 9, pp. 1347–55, 2014.

[59] B. Malavaud, O. Cussenot, N. Mottet, F. Rozet, A. Ruffion, L. Smets, and H. Stoevelaar, "Impact of adoption of a decision algorithm

including pca3 for repeat biopsy on the costs for prostate cancer diagnosis in france," *Journal of Medical Economics*, vol. 16, no. 3, pp. 358–63, 2013.

[60] A. Haese, A. de la Taille, H. van Poppel, M. Marberger, A. Stenzl, P. F. Mulders, H. Huland, C. C. Abbou, M. Remzi, M. Tinzl, S. Feyerabend, A. B. Stillebroer, M. P. van Gils, and J. A. Schalken, "Clinical utility of the pca3 urine assay in european men scheduled for repeat biopsy," *European Urology*, vol. 54, no. 5, pp. 1081–8, 2008.

[61] E. D. Crawford, K. O. Rove, E. J. Trabulsi, J. Qian, K. P. Drewnowska, J. C. Kaminetsky, T. K. Huisman, M. L. Bilowus, S. J. Freedman, J. Glover, W. L., and D. G. Bostwick, "Diagnostic performance of pca3 to detect prostate cancer in men with increased prostate specific antigen: a prospective study of 1,962 cases," *Journal of Urology*, vol. 188, no. 5, pp. 1726–31, 2012.

[62] A. de la Taille, J. Irani, M. Graefen, F. Chun, T. de Reijke, P. Kil, P. Gontero, A. Mottaz, and A. Haese, "Clinical evaluation of the pca3 assay in guiding initial biopsy decisions," *Journal of Urology*, vol. 185, no. 6, pp. 2119–25, 2011.

[63] B. Tombal, F. Ameye, A. de la Taille, T. de Reijke, P. Gontero, A. Haese, P. Kil, P. Perrin, M. Remzi, J. Schroder, M. Speakman, A. Volpe, B. Meesen, and H. Stoevelaar, "Biopsy and treatment decisions in the initial management of prostate cancer and the role of pca3; a systematic analysis of expert opinion," *World Journal of Urology*, vol. 30, no. 2, pp. 251–6, 2012.

[64] K. J. Wojno, F. J. Costa, R. J. Cornell, J. D. Small, E. Pasin, W. Van Criekinge, J. W. Bigley, and L. Van Neste, "Reduced rate of repeated prostate biopsies observed in confirmmdx clinical utility field study," *American Health & Drug Benefits*, vol. 7, no. 3, pp. 129–34, 2014.

[65] P. F. Pinsky, E. D. Crawford, B. S. Kramer, G. L. Andriole, E. P. Gelmann, R. Grubb, R. Greenlee, and J. K. Gohagan, "Repeat prostate biopsy in the prostate, lung, colorectal and ovarian cancer screening trial." *BJU International*, vol. 99, no. 4, pp. 775–9, 2007.

[66] 4Kscore®, "Announcing a drop in the price of the 4kscore™ test," http://4kscore.com/news/announcing-drop-price-4kscore-test/, may 2017, accessed: 2017-06-02.

[67] Gamma-Dynacare Medical Laboratories, "PCA3 Urine Test for Prostate Cancer," https://www.dynacare.ca/DYN/media/DYN/Pdf/Next/PCA3-Provider-Sheet.pdf, may 2012, accessed: 2017-06-02.

[68] F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa *et al.*, "Screening and prostate-cancer mortality in a randomized european study," *N Engl j Med*, vol. 2009, no. 360, pp. 1320–1328, 2009.

A Data Extraction

Data Extraction form

Name of reviewer:
Date:
Study ID (first author, year):
Notes:

Article's title:
Type of study
1. Randomized, controlled clinical trial / survey / observational …? 2. Was the study designed to evaluate the clinical utility of the new prognostic test, or was it a secondary analysis of data collected for other purposes? 3. Funding source: 4. To what phase does the study belong: Phase 1- screening /Phase 2- after positive biopsy/Phase 3- after negative biopsy/ Phase 4- add treatment 5. What was the testing scenario? (Is there in the study a comparison between cases and controls? Did the 2 groups have similar characteristics?)
Study population
6. Country and year where the study was conducted: 7. Number of subjects enrolled: 8. Number of subjects completed the study: 9. Duration of study: 10. Characteristics of subjects: 11. Inclusion/exclusion criteria: 12. Were the patients selected by the physician or randomly assigned?

13. Did the sample represent patients that would be tested in clinical practice?

14. Is there a concern for selection bias (systematic differences between baseline characteristics of the groups that are compared), explain:

Intervention

15. What is the used intervention?

16. Were investigators/physicians blinded to the test results?
(when they gave their first recommendation they didn't know about test results)

Outcome

17. What are the outcomes studied?

Outcome assessment

18. How were the outcomes assessed?

Results

B Quality Assessment

Rector et al checklist.

Supplementary Table – Quality assessment form

Name of reviewer:
Date:
Study ID (first author, year):
Notes:

Scoring procedure: 0 if ‘‘not clear’’ or ‘‘not a relevant item’’ or “not good quality”, 1 for ‘‘good quality’’, 2 for ‘‘excellent quality’’.

Article’s title:	**Score**
Type of study	
1. Was the study designed to evaluate the clinical utility of the new prognostic test, or was it a secondary analysis of data collected for other purposes? 2. Were there conflicts of interest? (0 if there is a conflict/ 2 if no conflict)	
Study population	
3. Was the clinical population clearly described including inclusion and exclusion criteria and subject participation? 4. How were the patients assigned to the chosen intervention? Where they selected by the physician or randomly assigned? 5. Is there a concern for selection bias (systematic differences between baseline characteristics of the groups that are compared)? 6. Did the sample represent patients that would be tested in clinical practice?	
Intervention	
7. Were the prognostic tests clearly described and conducted using a standardized, reliable, and valid method? 8. Was the test used and interpreted the same way by all sites/studies including any indeterminate test results? (Did they do and interpret as standards/ manufacturer says) 9. Were investigators blinded to the test results?	

Outcome	
10. Was the outcome being predicted clearly defined?	
Outcome assessment	
11. Was the outcome being predicted ascertained using a standardized, reliable, and valid method? (For example if there is a change in treatment, did a third party assess the change?) 12. Did everyone in the samples have a common starting point for follow up with respect to the outcome of interest including any treatments that could affect the outcome being predicted? (Did the patients receive any treatment /intervention that could affect the results/outcomes "DRE timing, 5-α-reductase inhibitors …/Were all the patients from same phase or were the patients who did the test from low risk group and who didn't from high risk group → overestimation since high risk groups are less likely to change treatment) 13. Is there a concern for performance bias (systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest)? 14. Is there a concern for detection bias (systematic differences between groups in how outcomes are determined)?	
Follow-up	
15. How complete was the follow up of subjects, and were losses to follow up related to the test results or the outcome being predicted? Was the duration of follow up adequate? (If no follow up in a study where there is no need for follow up, then the answer is 0 not applicable) 16. Is there a concern for attrition bias (systematic differences between groups in withdrawals from a study, e.g. data not available or exclusions)?	
17. Is there a concern for reporting bias (systematic differences between reported and unreported findings, e.g. are both significant and non-significant differences reported)?	

C 5 year results, screening tests

Table C.1: Results screening tests 5 years. Base case and sensitivity analysis

Variable and variations	*4Kscore®*		*PHI*	
	Cost	QALY	Cost	QALY
Base case	$6,000	4.38	$7,839	4.33
Sensitivity Analysis				
Variation of discount rate				
No discount rate	$6,537	4.80	$8,690	4.75
3% discount rate	$6,195	4.54	$8,165	4.49
10% discount rate	$5,528	4.02	$7,138	3.98
Variation of AS rate				
10%	$6,079	4.37	$7,936	4.32
20%	$5,894	4.38	$7,754	4.33
25%	$5,797	4.38	$7,661	4.33
40%	$5,518	4.39	$7,386	4.34
50%	$5,332	4.39	$7,206	4.35
Variation of probability of recurrence in intermediate-high-risk group				
2.9%	$5,923	4.38	$7,707	4.33
5.3%	$6,023	4.37	$7,912	4.33
6.5%	$6,066	4.37	$8,013	4.32
7.7%	$6,115	4.37	$8,117	4.32
Variation of utilities				
0.95; 0.95; 0.81; 0.48	$5,987	4.41	$7,845	4.38
0.89; 0.89; 0.75; 0.42	$5,987	4.34	$7,845	4.28
0.85; 0.85; 0.70; 0.40	$5,987	4.30	$7,845	4.21
1.00; 1.00; 1.00; 1.00	$5,987	4.47	$7,845	4.48

Variation of Follow Up cost				
$200.0	$5,789	4.38	$7,700	4.33
$232.5	$5,863	4.38	$7,754	4.33
$395.0	$6,233	4.38	$8,025	4,33
Variation of initial positive biopsy probability				
33.0%	$5,669	4.39		
56.5%	$6,386	4.36		
65.0%	$6,656	4.35		
Variation of rate of initial biopsies				
40.0%	$5,679	4.38		
63.8%	$6,519	4.36		
66.25%			$8,400	4.31
84.50%			$8,995	4.28
Variation of test cost				
$410			$8,105	4.33
$670			$8,365	4.33
Variation of false negative rate				
10.4%			$5,661	4.36
19.8%			$6,858	4.34
Variation of cut-off, FN = false negative				
$27.6 (performed biopsies 84.5% and FN 26%)			$8,771	4.29
$40.3 (performed biopsies 44.3% and FN 28.4%)			$7,672	4.33
$50.9 (performed biopsies 25.5% and FN 33.1%)			$7,553	4.35

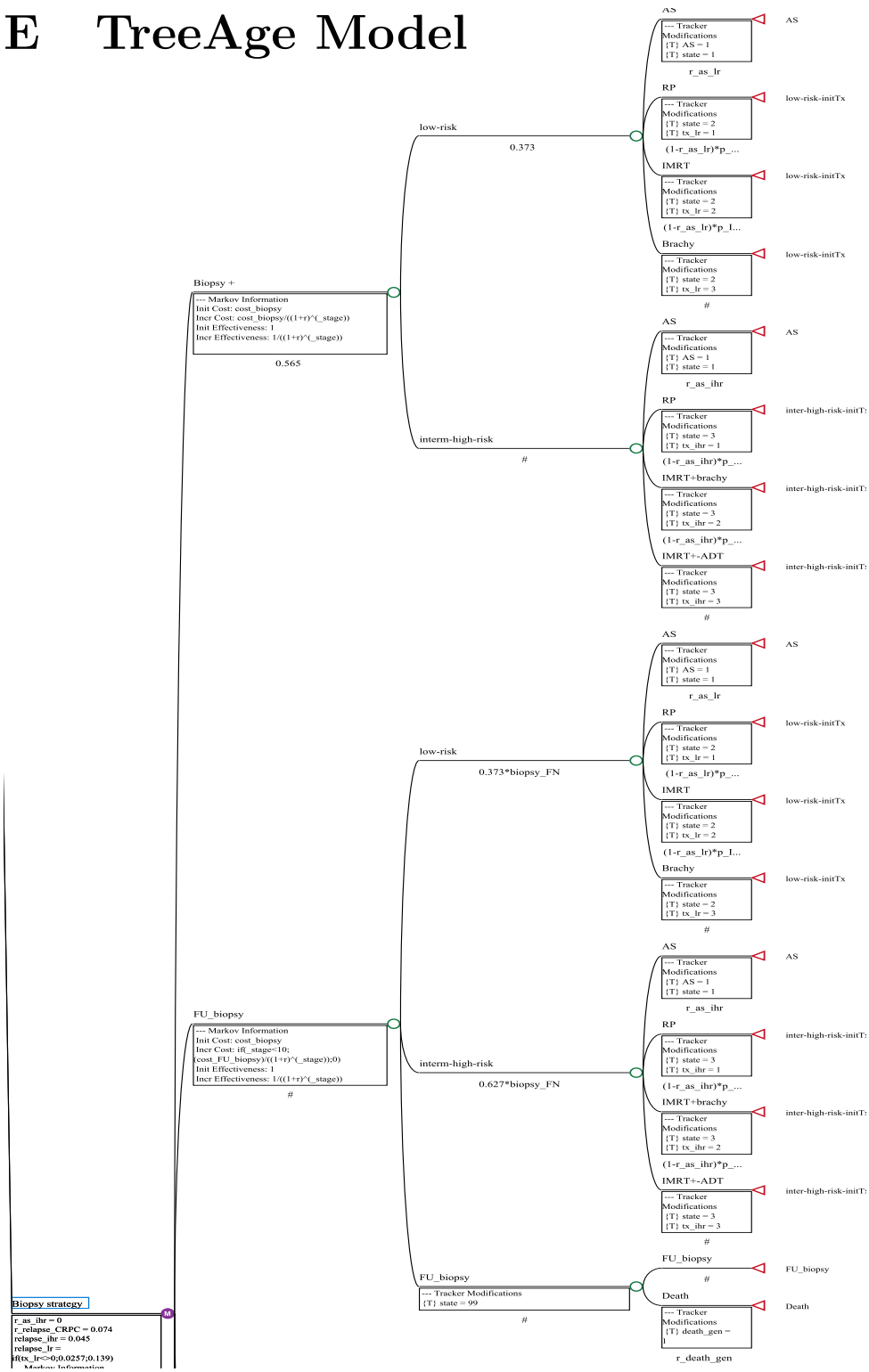
D 5 year results, tests after a negative biopsy

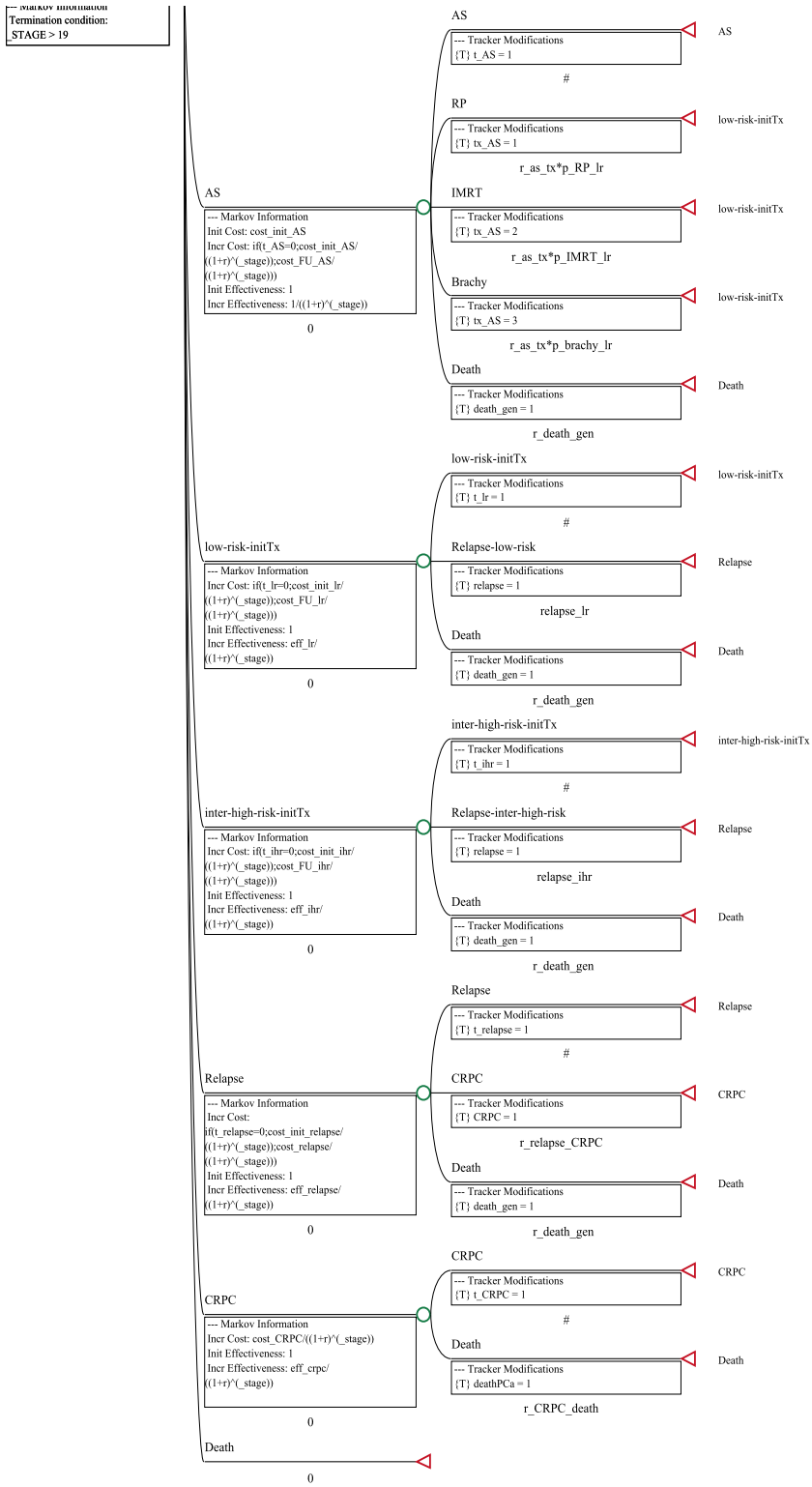
Table D.1: Results tests after negative biopsy 5 years. Base case and sensitivity analysis

Variable and variations	*PCA3*		*ConfirmMDx*	
	Cost	QALY	Cost	QALY
Base case	$8,028	4.29	$8,246	4.29
Sensitivity Analysis				
Variation of discount rate				
No discount rate	$8,643	4.71	$8,876	4.71
3% discount rate	$8,262	4.45	$8,490	4.45
10% discount rate	$7,500	3.94	$7,717	3.94
Variation of AS rate				
10%	$8,119	4.29	$8,343	4.29
20%	$7,947	4.29	$8,168	4.29
25%	$7,865	4.30	$8,085	4.30
40%	$7,608	4.30	$7,830	4.30
50%	$7,439	4.31	$7,664	4.31
Variation of probability of recurrence in intermediate-high-risk group				
2.9%	$7,827	4.29	$8,047	4.29
5.3%	$8,133	4.29	$8,047	4.29
6.5%	$8,288	4.28	$8,500	4.28
7.7%	$8,436	4.28	$8,644	4.28
Variation of utilities				
0.95; 0.95; 0.81; 0.48	$8,027	4.35	$8,252	4.35
0.89; 0.89; 0.75; 0.42	$8,027	4.23	$8,252	4,23

0.85; 0.85; 0.70; 0.40	\$8,027	4.15	\$8,252	4.15
1.00; 1.00; 1.00; 1.00	\$8,027	4.47	\$8,252	4.47
Variation of FU cost				
\$200.0	\$7,920	4.29	\$8,143	4.29
\$232.5	\$7,960	4.29	\$8,183	4.29
\$395.0	\$8,160	4.29	\$8,387	4.29
Variation of repeat biopsy rate				
37%	\$8,004	4.29		
60%	\$8,051	4.29		
Variation of test cost				
\$2,220			\$8,097	4.29
\$3,330			\$8,174	4.29

E TreeAge Model




--- Markov Information
Termination condition:
_STAGE > 19
AS
--- Markov Information
Init Cost: cost_init_AS
Incr Cost: if(t_AS=0;cost_init_AS/
((1+r)^(_stage));cost_FU_AS/
((1+r)^(_stage)))
Init Effectiveness: 1
Incr Effectiveness: 1/((1+r)^(_stage))
0
AS
--- Tracker Modifications
{T} t_AS = 1
#
AS
RP
--- Tracker Modifications
{T} tx_AS = 1
r_as_tx*p_RP_lr
low-risk-initTx
IMRT
--- Tracker Modifications
{T} tx_AS = 2
r_as_tx*p_IMRT_lr
low-risk-initTx
Brachy
--- Tracker Modifications
{T} tx_AS = 3
r_as_tx*p_brachy_lr
low-risk-initTx
Death
--- Tracker Modifications
{T} death_gen = 1
r_death_gen
Death
low-risk-initTx
--- Markov Information
Incr Cost: if(t_lr=0;cost_init_lr/
((1+r)^(_stage));cost_FU_lr/
((1+r)^(_stage)))
Init Effectiveness: 1
Incr Effectiveness: eff_lr/
((1+r)^(_stage))
0
low-risk-initTx
--- Tracker Modifications
{T} t_lr = 1
#
low-risk-initTx
Relapse-low-risk
--- Tracker Modifications
{T} relapse = 1
relapse_lr
Relapse
Death
--- Tracker Modifications
{T} death_gen = 1
r_death_gen
Death
inter-high-risk-initTx
--- Markov Information
Incr Cost: if(t_ihr=0;cost_init_ihr/
((1+r)^(_stage));cost_FU_ihr/
((1+r)^(_stage)))
Init Effectiveness: 1
Incr Effectiveness: eff_ihr/
((1+r)^(_stage))
0
inter-high-risk-initTx
--- Tracker Modifications
{T} t_ihr = 1
#
inter-high-risk-initTx
Relapse-inter-high-risk
--- Tracker Modifications
{T} relapse = 1
relapse_ihr
Relapse
Death
--- Tracker Modifications
{T} death_gen = 1
r_death_gen
Death
Relapse
--- Markov Information
Incr Cost:
if(t_relapse=0;cost_init_relapse/
((1+r)^(_stage));cost_relapse/
((1+r)^(_stage)))
Init Effectiveness: 1
Incr Effectiveness: eff_relapse/
((1+r)^(_stage))
0
Relapse
--- Tracker Modifications
{T} t_relapse = 1
#
Relapse
CRPC
--- Tracker Modifications
{T} CRPC = 1
r_relapse_CRPC
CRPC
Death
--- Tracker Modifications
{T} death_gen = 1
r_death_gen
Death
CRPC
--- Markov Information
Incr Cost: cost_CRPC/((1+r)^(_stage))
Init Effectiveness: 1
Incr Effectiveness: eff_crpc/
((1+r)^(_stage))
0
CRPC
--- Tracker Modifications
{T} t_CRPC = 1
#
CRPC
Death
--- Tracker Modifications
{T} deathPCa = 1
r_CRPC_death
Death
Death
0

F Rates from previous study

Table F.1: Base rates extracted from previous study [6]

Variable	Value
TRUSGB base rates	
Rate of positive TRUSGB in case of clinical suspicion of prostate cancer	56.5%
Rate of significant prostate cancer among positive biopsies	62.7%
Treatment Allocation	
Rate of low-risk prostate cancer on AS	15.0%
Rate of delayed treatment following AS at 5 years	25.0%
Rate of low-risk prostate cancer undergoing RP	35.0%
Rate of low-risk prostate cancer undergoing brachytherapy	15.0%
Rate of intermediate/high-risk prostate cancer undergoing RP	30.0%
Rate of intermediate/high-risk prostate cancer undergoing RT	30.0%
Rate of intermediate/high-risk prostate cancer undergoing brachytherapy + EBRT	10.0%
Rate of intermediate/high-risk prostate cancer undergoing RT + ADT	30.0%
Recurrences and survivals	
5-year biochemical recurrence in AS patients who underwent treatment	53.0%
5-year biochemical recurrence in overall treated cohort of low risk	13.0%
7-year biochemical recurrence after brachytherapy + EBRT + ADT in high risk	43.0%

7-year biochemical recurrence after brachytherapy + EBRT in intermediate risk	10.0%
7-year biochemical recurrence after EBRT in intermediate risk	18.6%
Median time elapsed from biochemical recurrence (from ADT initiation) to CRPC	9.1 years
Median time from CRPC to death from prostate cancer) to CRPC	25 months
Costs	
Biopsy	$650
1-year FU when biopsy negative	$287
AS (first year)	$745
1-year FU in AS	$353
Treatment in low-risk group	$8,860
Treatment in intermediate/high-risk group	$10,233
1-Year FU remission in low-risk group	$141
1-Year FU remission in intermediate/high-risk group	$164
First year relapse	$8,562
Relapse in second year to CRPC phase	$4,640
1-Year in CRPC phase	$4,640

G 10, 15 and 20 year results from previous study

Table G.1: Incremental cost and QALY summary table using TRUSGB, results from previous study

Variable and variations	Incremental Cost (10 years)	Incremental QALY (10 years)	Incremental cost (15 years)	Incremental QALY (15 years)	Incremental Cost (20 years)	Incremental QALY (20 years)
Base Case	$11,526	7.22	$14,954	9.11	$17,495	10.19
Sensitivity Analysis						
Variation of discount rate						
No discount rate	$13,618	8.81	$19,713	12.16	$25,473	14.59
3% discount rate	$12,285	7.79	$16,560	10.15	$20,056	11.63
10% discount rate	$10,025	6.08	$11,979	7.17	$13,138	7.66
Variation of AS rate						
10%	$11,616	7.21	$15,027	9.09	$17,582	10.17
20%	$11,456	7.23	$14,862	9,11	$17,403	10.18
25%	$11,375	7.23	$14,774	9.12	$17,285	10.20

Variation of probability of recurrence in intermediate-high-risk group						
2.9%	$10,615	7.25	$13,321	9.17	$15,413	10.28
5.3%	$11,954	7.21	$15,651	9.07	$18,380	10.13
6.5%	$12,568	7.19	$16,709	9.03	$19,634	10.06
7.7%	$13,157	7.17	$17,650	8.99	$20,793	10.01
Variation of utilities						
0.95; 0.95; 0.81; 0.48	$11,535	7.34	$14,932	9.26	$17,461	10.36
0.89; 0.89; 0.75; 0.42	$11,535	7.10	$14,932	8.94	$17,461	9.99
0.85; 0.85; 0.70; 0.40	$11,535	6.93	$14,932	8.71	$17,461	9.73
1.00; 1.00; 1.00; 1.00	$11,535	7.64	$14,932	9.71	$17,461	10.90

Table G.2: Incremental cost and QALY summary table using MRGB, results from previous study

Variable and variations	Incremental Cost (10 years)	Incremental QALY (10 years)	Incremental cost (15 years)	Incremental QALY (15 years)	Incremental Cost (20 years)	Incremental QALY (20 years)
Base Case	$10,011	7.31	$12,814	9.25	$14,866	10.37
Sensitivity Analysis						
Variation of discount rate						
No discount rate	$11,652	8.92	$16,682	12.37	$21,303	14.91
3% discount rate	$10,593	7.89	$14,121	10.32	$16,927	11.86
10% discount rate	$8,802	6.15	$10,414	7.27	$11,345	7.78
Variation of AS rate						
10%	$10,006	7.31	$12,812	9.24	$14,853	10.36
20%	$9,983	7.31	$12,783	9.25	$14,842	10.37
25%	$9,974	7.31	$12,781	9.25	$14,838	10.37
40%	$9,944	7.31	$12,742	9.24	$14,758	10.37
50%	$9,924	7.31	$12,723	9.25	$14,750	10.36
Variation of probability of recurrence in intermediate-high-risk group						
2.9%	$9,018	7.33	$11,051	9.30	$12,553	10.46

5.3%	$10,473	7.29	$13,628	9.20	$15,894	10.30
6.5%	$11,159	7.27	$14,723	9.16	$17,194	10.24
7.7%	$11,805	7.25	$15,733	9.11	$18,427	10.17
Variation of utilities						
0.95; 0.95; 0.81; 0.48	$9,997	7.40	$12,801	9.37	$14,831	10.50
0.89; 0.89; 0.75; 0.42	$9,997	7.21	$12,801	9.12	$14,831	10.22
0.85; 0.85; 0.70; 0.40	$9,997	7.08	$12,801	8.95	$14,831	10.03
1.00; 1.00; 1.00; 1.00	$9,997	7.63	$12,801	9.72	$14,831	10.93

H 5 year results from previous study

Table H.1: Results 5 years from previous study. Base case and sensitivity analysis

Variable and variations	*TRUSGB*		*MRGB*	
	Cost	QALY	Cost	QALY
Base case	$8,032	4.29	$7,124	4.32
Sensitivity Analysis				
Variation of discount rate				
No discount rate	$8,654	4.70	$7,583	4.74
3% discount rate	$8,270	4.44	$7,301	4.48
10% discount rate	$7,505	3.94	$6,735	3.97
Variation of AS rate				
10%	$8,122	4.28	$7,136	4.32
20%	$7,947	4.29	$7,117	4.32
25%	$7,868	4.29	$7,106	4.32
40%			$7,075	4.32
50%			$7,056	4.32
Variation of probability of recurrence in intermediate-high-risk group				
2.9%	$7,827	4.29	$6,901	4.32
5.3%	$8,134	4.28	$7,244	4.32
6.5%	$8,284	4.28	$7,411	4.31
7.7%	$8,428	4.28	$7,580	4.31
Variation of utilities				
0.95; 0.95; 0.81; 0.48	$8,035	4.35	$7,127	4.37
0.89; 0.89; 0.75; 0.42	$8,035	4.22	$7,127	4.27
0.85; 0.85; 0.70; 0.40	$8,035	4.14	$7,127	4.21

1.00; 1.00; 1.00; 1.00	$8,035	4.47	$7,127	4.46