Clinical Prediction Models for Prostate Cancer

from development to validation and implementation

Johannes Franciscus Marcus Verbeek

ISBN 978-94-6380-916-0
Cover Jane Klein, OptimaForma.net, Nijmegen
Print Proefschriftmaken.nl

The printing of this thesis was financially supported by :
Stichting Wetenschappelijk Onderzoek Prostaatkanker
Stichting Urologisch Wetenschappelijk Onderzoek
Erasmus University Medical Center Rotterdam

Clinical Prediction Models for Prostate Cancer

from development to validation and implementation

Klinische voorspellingsmodellen voor prostaatkanker
van ontwikkeling tot validatie en implementatie

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam

op gezag van de Rector Magnificus
Prof. Dr. R.C.M.E. Engels
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
10 september 2020 om 11.30 uur precies

door

Johannes Franciscus Marcus Verbeek
geboren te Nijmegen

Erasmus University Rotterdam

PROMOTIECOMMISIE

Promotoren

Prof. Dr. M.J. Roobol-Bouts

Prof. Dr. E.W. Steyerberg

Overige leden

Prof. Dr. P.J.E. Bindels

Prof. Dr. S. le Cessie

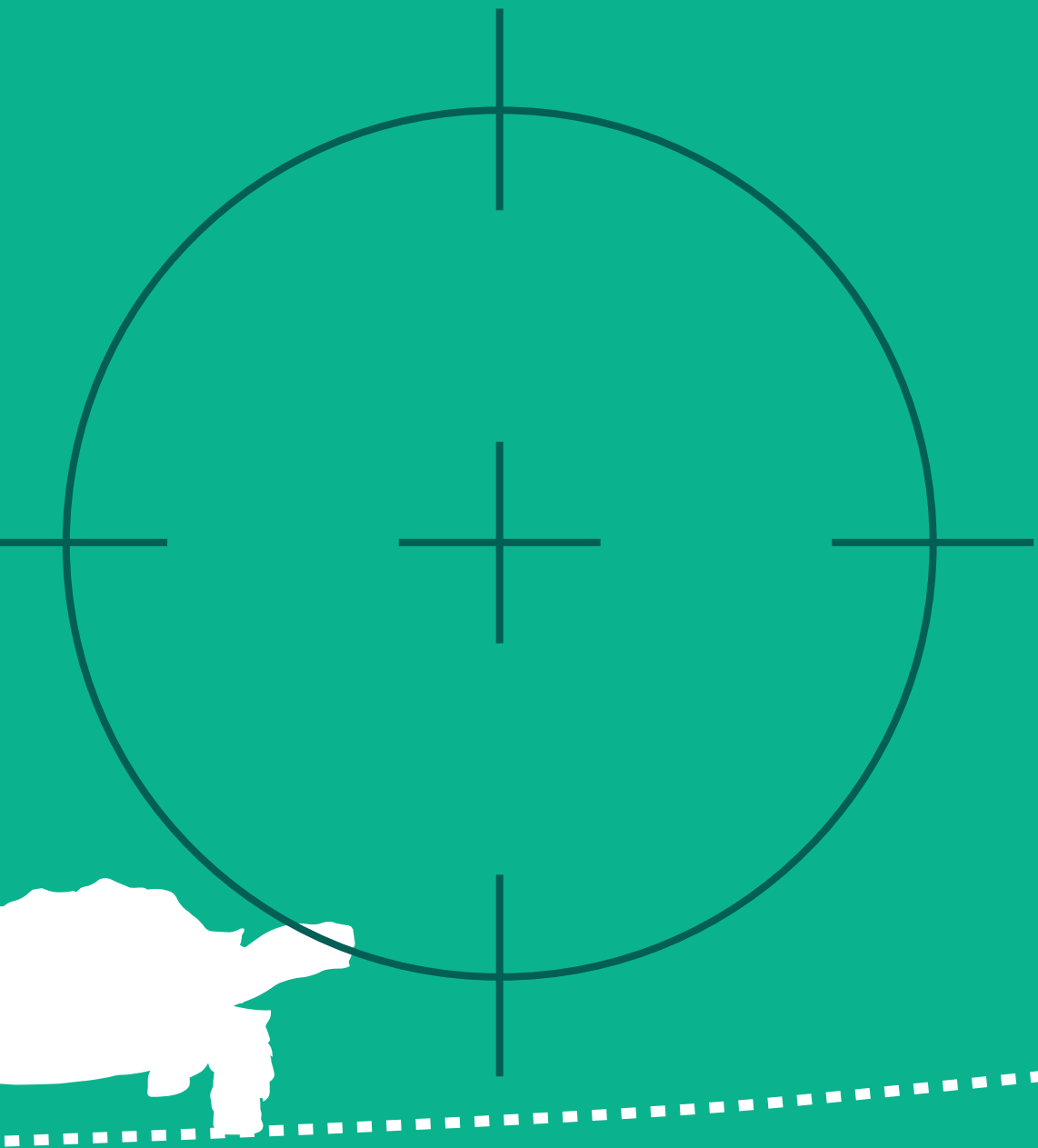
Prof. Dr. C.H. Bangma

Paranimfen

Drs. F.H. Drost

Drs. D.F. Osses

TABLE OF CONTENTS



CHAPTER 1

General introduction and outline of this thesis

1.1 INTRODUCTION

Traditionally, medicine has been about diagnosis, treatment and, to a lesser extent, prevention. In the last 50 years, screening programs have become common 'medical practice'. These programs aim to catch disease at an early, preferably asymptomatic, state to enhance cure rate. There are basically three components to consider with respect to the evidence for effectiveness of a screening program: Proven mortality reduction, proven quality of life benefit based on quality-adjusted life years and cost-effectiveness. If hard evidence, i.e. quantitative information from well-designed studies, of any of these parts of the screening chain is not available, the discussion on screening acceptability remains. To ensure a balanced discussion, we should gather key information on those who test positive, the positive predictive value, duration of the preclinical detectable phase, differences in cure rate or case fatality rate, false positive and negative rates, and overdiagnosis. Prediction models may help to refine these programs by considering the risks of individual patients rather than risks of populations defined by age only.

In this thesis, I discuss the benefits and harms of prostate-specific antigen (PSA) testing for the early detection of prostate cancer. These are classically calculated at the population level. In this thesis, I aim to determine whether quantitative prostate cancer risk information — preferably designed as a simple predictive device or tool — can add value when making health care decisions to screen for prostate cancer and/or active surveillance of diagnosed patients.

The prostate gland, hyperplasia and cancer

Considerations on the early detection of prostate cancer (PCa) is based on basic knowledge of the function, histology and cancer occurrence in the prostate gland, which is summarized here. The prostate gland is a conical body, classically described as "walnut-shaped", surrounding the urethra near the bottom of the bladder [1]. The prostate has two major functions: firstly, physically, through its mass and musculature, it controls urine output from the bladder and the transmission of seminal fluid during ejaculation; secondly, as an exocrine gland, it contributes to the fluid that protects the semen, in addition to the fluid from the seminal vesicles [2]. Histologically, the prostate can be divided into four zones: peripheral, transition, central, and anterior fibromuscular. Prostatic diseases show a zonal distribution with approximately 70% of adenocarcinoma tumors arising in the peripheral zone, while benign prostatic hyperplasia (BPH) most frequently occurs in the transition zone [3]. Hyperplasia in BPH refers to an increase of the number of cells in the transition zone of the prostate, thereby compressing the urethra and leading to lower urinary tract symptoms. BPH is a common prostatic disorder, with prevalence increasing with age, from 50% at age 50 up to 80% at age 70 [4]. PCa develops when DNA mutations cause uncontrolled cell growth.

Most PCa are small, often grow slowly, and are compactly localized in the prostate. Therefore, they are mostly asymptomatic, and may never cause any health problems.

One in 8-11 men will be diagnosed with PCa [5]. As with benign prostatic hyperplasia, older men are more at risk of developing PCa, especially those with a positive family history and those who are African American. In the Netherlands, PCa incidence continues to increase, with in 2018 an estimated 12,646 cases [6]; in 2017, PCa was responsible for 4% of all male deaths, a total of 2862. In the last two decades, PCa mortality has declined, while incidence has increased (Figure 1).

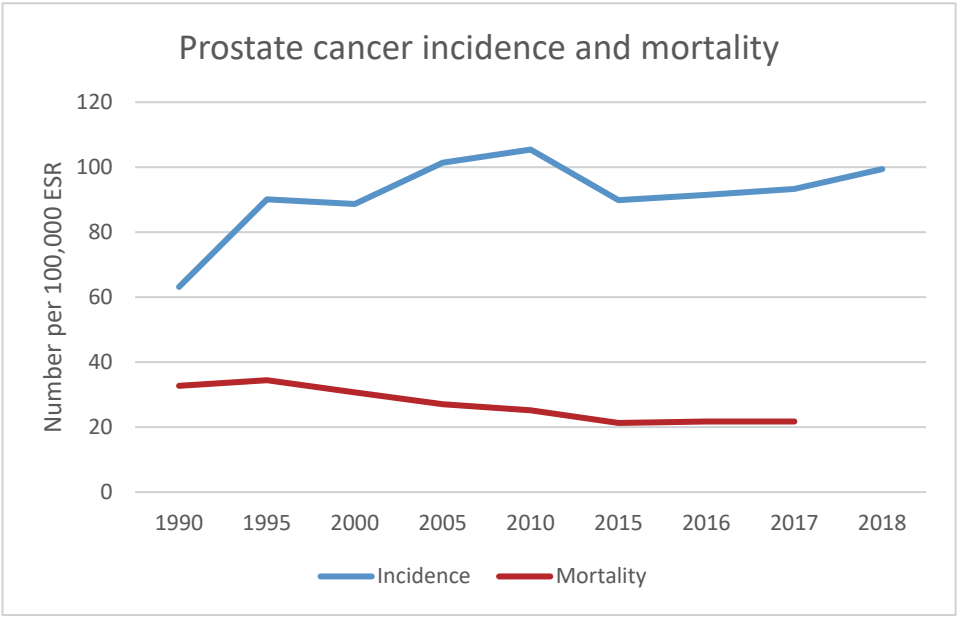


Figure 1. Prostate cancer incidence and mortality in Dutch men (1990-2018). Cancer rates are defined by the number of cases per 100,000 person-years and age-standardized to the European standard population (ESR).

These data suggest a high PCa survival rate, however, as the disease develops slowly, death may be due to other causes before PCa is clinically advanced. This is also indicated by autopsy series, where PCa is detected in approximately 30% of men aged 55 and in approximately 60% of men aged 80 [7]. PCa survival is related to the extent of the tumor at time of diagnosis. For localized disease where the cancer is confined to the prostate, five-year relative survival is close to 100% compared to 30% among those diagnosed with distant metastases [8]. Thus, a screening program that accurately identifies asymptomatic men with aggressive but still localized tumors may substantially reduce prostate cancer morbidity, painful metastases, and mortality.

Prostate-specific antigen

Most PCa are asymptomatic and are therefore difficult to identify at early stages. Prostate-specific antigen (PSA) has revolutionized early PCa detection. PSA is a 35 kiloDalton serine protease of the kallikrein family of proteins, produced by prostate epithelial cells. It was initially found in prostatic tissue by Dr. Rubin Hyman Flocks in 1960, and in seminal fluid by Dr. Mitsuwo Hara in 1964 [9]. PSA levels may be elevated in men with PCa as PSA production is increased by tumor cells and because tissue barriers between the prostate gland lumen and the capillary are disrupted, releasing more PSA into the serum. However, PSA can also be elevated in benign conditions, particularly BPH and prostatitis. Worldwide, PSA is the most widely used marker in diagnosis and follow-up of any cancer. PSA was first introduced to detect PCa recurrence and disease progression after treatment, but in the early 1990s, it became a standard for PCa early detection [10].

PCa is occasionally palpable by digital rectal examination (DRE) as it causes hard or irregular nodules. Currently, the only other way to detect and confirm PCa is through a prostate biopsy by a urologist using transrectal ultrasonography (TRUS). The ultrasound probe is inserted in the rectum and biopsies can be performed though the rectal wall or perineum. Up to 10-12 biopsies are systematically taken; additional biopsies can be performed if suspicious lesions are found on TRUS or magnetic resonance imaging (MRI). Once the histological data have been gathered, the pathologist evaluates the tissue, grading PCa using the International Society of Urological Pathology (ISUP) consensus on Gleason score for PCa [11]. This is a validated alternative of the Gleason score and includes five distinct grade groups based on modified Gleason score groups: Grade group 1 = Gleason score $\leq 3+3$, Grade group 2 = Gleason score $3 + 4 = 7$, Grade group 3 = Gleason score $4 + 3 = 7$, Grade group 4 = Gleason score 8, and Grade group 5 = Gleason scores 9 and 10. ISUP grade 1, previously known as Gleason score 3 + 3 and ISUP grade 2 (Gleason score 3+4) in the absence of secondary growth patterns like cribriform growth and intraductal carcinoma, are considered low-risk PCa [12].

1.2 PROSTATE CANCER SCREENING

Screening programs should reduce disease-specific morbidity and mortality, improve quality of life and be cost-effective. A good screening test is minimally invasive, easily performed, objective, and acceptable to the general population. The prostate-specific antigen (PSA) test matches these characteristics. However, PCa screening with PSA has been hotly debated for more than three decades, as demonstrated by the many different recommendations and guidelines [10, 13, 14]. Ideally, screening should reduce numbers of PCa deaths without

excessively harming screened people. The World Health Organization state that screening is only permitted if a net benefit to those screened can be established [15].

Relevant endpoint of screening

Efficacy studies preferably apply randomized controlled designs. Randomized screening studies that assess the value of diagnostic PCa screening tools include large numbers of participants and a long follow-up period. When PSA testing was introduced in the 1990s, it was deemed opportunistic as it was not based on any protocol [16]. Later on, two major screening research trials were conducted: the European Randomized Study of Screening for Prostate Cancer (ERSPC) trial, and the American Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial [17, 18]. The introduction of PSA and the screening trials resulted in a concomitant increase in PCa incidence and a decline in PCa mortality. However, to what extent PSA screening reduced absolute mortality remains unknown [19]. In addition, changes in patient management such as improvements in surgery, radiation therapy, and treatment of metastatic disease have also contributed to lowering mortality rates in the same period [20].

The ERSPC trial showed a reduction in PCa mortality following PCa screening with PSA. This randomized trial included a total of 182,160 men aged 50–74 from eight centers across Europe. Of these, 162,388 men were in the target age group (55-69). Men who entered the screening arm received PSA testing at an interval of 2-4 years and systematically received TRUS prostate biopsy with 6 cores (sextant) if PSA was elevated >3.0 ng/ml. Men in the control group received standard care. After a 13-year median follow-up period, 10.2% PCa was found in the screening arm vs 6.8% in the control group, with a PCa death of 0.49% vs 0.61% respectively. After even longer follow-up (16 years) and after adjustment for nonparticipation, the difference in absolute PCa mortality was 0.18% at 16 years [21]. In other words, for every 570 men invited to screening, 18 have to be diagnosed and treated to prevent 1 PCa death. It should be noted that these numbers to screen differ due to differences in the screening program at each center: from 65 in Switzerland to 7 in Sweden. In relative terms, an individual can reduce prostate cancer mortality risk by up to 50% by complying with the screening program. However, the absolute numbers of benefit in the population are small [22] and no difference in overall mortality was found [23]. In contrast, the PLCO screening trial reported no PCa mortality benefit, even after 15 years of follow-up [24]. A major limitation of this trial is the high PSA testing rate in the control arm which means that the comparison between organized versus opportunistic screening can be discussed.

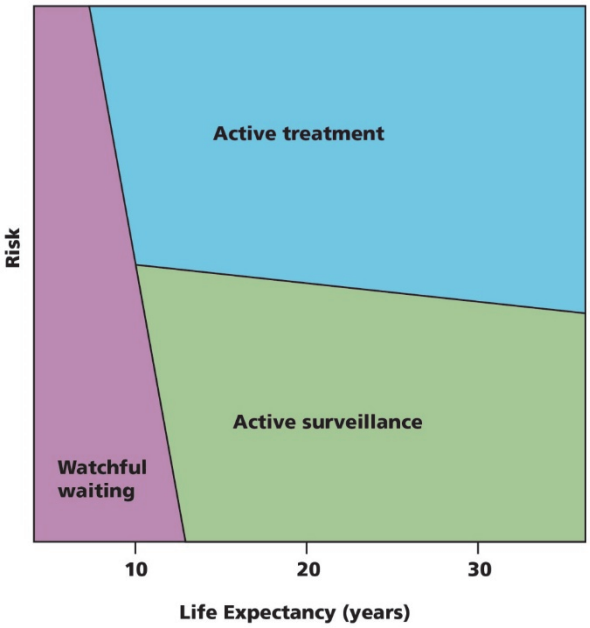
The ERSPC provides evidence for gains from PCa screening, however, screening is also associated with false-positive results, biopsy complications, and overdiagnosis. The number of overdiagnosed PCa ranges from 2% up to 67% of all screen-detected prostate cancer [25].

Overdiagnosis of PCa generally leads to overtreatment, however, this can be reduced by active surveillance (AS); this is explained in more detail in the following section. The long-term effects of active treatment are unclear compared to AS and other conservative strategies, but are associated with sexual and urinary difficulties [26]. The Netherlands no screening program is available. Nevertheless, the U.S. Preventative Services Task Force, EAU and AUA guidelines recommend that men aged 55–69 years should be informed about the benefits and harms of screening, and offer PSA testing only after an informed choice has been made [14, 27, 28]. Testing for PCa is a complex individual-based decision-making process, taking into consideration competing risks from other co-morbidities when estimating a survival benefit from the early detection of clinically significant PCa. Important questions are: 1) does the patient have PCa, 2) if so, what is the natural history of the disease, and 3) will the patient benefit from early detection and treatment of PCa? As this decision-making process is complex, the identification and weighing of important predictive variables are required. Today's challenge is to combine novel biomarkers and modern technology into a risk profile to optimize this decision-making process for individual health care and public health. This can be constructed using prediction models which aim to improve patient selection on the one hand, while simultaneously reducing overdiagnosis and overtreatment on the other [29].

1.3 ACTIVE SURVEILLANCE

Based on and summarized from: Verbeek JFM, Roobol MJ, Steyerberg EW. Risk-based selection for active surveillance. In: L Klotz (Ed.). Active Surveillance for Localized Prostate Cancer – a new paradigm for clinical management (pp. 53-64). Springer International Publishing, 2018.

As reported by the ERSPC trial, PSA screening reduces PCa mortality [17]. However, the number of men to be screened to prevent one case of PCa from dying from PCa is substantial. The main drawback of PSA-guided PCa screening is overdiagnosis. Overdiagnosis occurs when PCa is detected by screening, while the cancer has no impact on survival. It is more likely if patients have a shorter life expectancy due to their age or comorbidity, or if they have a high probability of dying from other causes than PCa. The detection of these clinically insignificant PCa, or those with a low risk of disease progression, subsequently leads to overtreatment and hence unnecessary adverse effects [30].



	Watchful Waiting	Active Surveillance	Active Treatment
Focus	managing symptoms	to delay/avoid AT	immediate treatment
Risk of PCa progression	very low – high in combination with short LE	low - intermediate	intermediate - high
Treatment intent	palliative	curative	curative
Follow-up	patient-specific	predefined schedule	after treatment
Life-expectancy	<10 years	>10 years	>10 years

Figure 2. The three different treatment strategies based on life-expectancy and risk of PCa progression. The WW line with increasing life expectancy illustrates the arbitrary nature of the 10-year cut off point; men with high risk of PCa progression with very short life expectancy are still suitable for WW, and those with a very low risk of PCa may also be suitable for WW. The downwards slope of the division line between Active Treatment (AT) and AS with increasing life expectancy illustrates that AT could be feasible for young men. (Adapted from Bruinsma S and Nieboer D, in L Klotz (Ed.), Active surveillance for localized prostate cancer, chapter 14; 2018, with permission).

Conservative strategies such as watchful waiting (WW) and active surveillance (AS) are adopted to reduce the harms of screening by reducing overtreatment of men with a low risk of PCa progression [25]. WW is commonly used for patients with a limited life expectancy; it is a palliative strategy with much less intense observation, followed by treatments for patients whose disease progresses. AS is a curative strategy for patients with a low risk of PCa progression, with the aim of avoiding overtreatment [31]. After an initial period of observation, AS is accompanied by rigorous and invasive follow-up, with delayed curative treatment for those with disease progression [32]. It is difficult to scientifically underpin which patients are eligible for AS, as we need to consider the risks of progression, life expectancy, and treatment effectiveness in a dynamic context. In Figure 2, the one-size-fits-all context for AS patient selection and follow-up is described.

Qualitative considerations on patient selection for active surveillance

The main goal of AS is to reduce overtreatment in patients with low-risk prostate cancer [33, 34]. Men with an initially low risk of PCa progression are usually considered to be suitable candidates for AS, provided they have a reasonable life expectancy, e.g., more than 10 years. They should be distinguished from those diagnosed with a progressive PCa who are more likely to die from PCa and who would substantially benefit from immediate active treatment [32, 35]. On the other hand, men should not be selected for AS if their lifetime risk of disease progression is low. In these cases, WW would be the optimal treatment option. The ProtectT trial compared radiotherapy, surgery and active monitoring [36]. Active monitoring is a variant between WW and AS. Findings from the recent ProtecT screening study suggest that two-thirds of patients diagnosed with PCa may be eligible for AS [18, 36, 37]. However, international AS cohorts differ in their inclusion criteria for AS patients; Table 1 [38-47].

Overall, current guidelines recommend patients as being most suitable for AS if they have pretreatment clinical stage T1(c) or T2a prostate cancer, serum PSA $<$10 ng/ml, a biopsy Gleason score of six, a maximum of two tumor-positive biopsy core samples and/or a maximum of 50% of cancer per core [31]. Some guidelines include statements that patients with stage T2b–T2c can also be recommended for AS. The Dutch Urology Association (DUA) guideline even recommends selecting patients with T3 for AS. Age and comorbidity are relevant, because a considerable life expectancy is important for AT, and hence AS, to show any long-term benefit. Finally, some guidelines state that patients' preferences should be considered in order to reduce the dropout rate of AS patients due to anxiety [31].

Table 1. Patient and biopsy-based inclusion criteria in different active surveillance protocols.

Active surveillance protocol	Clinical stage	PSA	Gleason score	Positive cores	Core positivity (%)	PSAD (ng/dL)
PRIAS [38]	≤T2	≤10	≤3+3	≤2	-	≤0.20
Sunnybrook [39]	-	≤10*	≤3+3*	-	-	-
Royal Marsden [48]	≤T2a	-**	≤3+3**	≤50%	-	-
Johns Hopkins [40]	T1c-T2a†	≤10†	≤3+3	≤2	≤50	<0.15
UCSF [41]	≤T2a	≤10	≤3+3	≤33%	≤50	≤0.15
UM [42]	≤T2a	≤10	≤3+3	≤2	≤20	-
UC [34]	≤T2a	≤10	≤3+3	≤3	<50	-
Australian [43]	≤T2a	<10	≤3+3	<20%	<30	-
Göteborg [44]	≤T2a	≤10	≤3+3	-	-	-
MSKCC [46]	≤T2a	≤10	≤3+3	≤3	≤50	-
Japan [47]	T1c	≤20	≤3+3	≤2	≤50	-

The grayed shaded blocks show similarities between the cohorts resulting in a more stringent inclusion of "low risk". The white shaded areas show a wider range of criteria, which allows for a slightly higher risk inclusion. *Includes Gleason ≤3+4 and / or PSA ≤ 20 if life expectancy ≤10yrs / or present comorbidities. ** Includes patients with Gleason ≤3+4 if age >65 year and PSA <15. † T2a only if PSA ≤10. PSAD: Prostate specific antigen density, PRIAS: Prostate Cancer Research International Active Surveillance, UCSF: University of California San Francisco, UM: University of Miami, UC: University of Copenhagen, MSKCC: Memorial Sloan Kettering Cancer Center New York.

Follow-up schedule for active surveillance

Just as there are many different inclusion criteria for AS, the follow-up schedule between cohorts is also different. They can be split in two major branches. The first is the Prostate cancer Research International: Active Surveillance (PRIAS) [49]. In the PRIAS protocol PSA is measured every 3 months and biopsies are taken at year 1, 4, 7 and 10 after an MRI has been performed, Figure 3. Once a patient is included for AS, after 3 months the patient receives an mpMRI and if a lesion is found, targeted biopsies are performed. This reduces the isclassification found at the beginning of AS [50]. The other branch performs a biopsy on a yearly schedule [51].

	Year	1					2				3		4		5		6		7	
	Month	*0*	*3*	*6*	*9*	*12*	*15*	*18*	*21*	*24*	*30*	*36*	*42*	*48*	*54*	*60*	*66*	*72*	*78*	*84*
PRIAS-study	*PSA*	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	DRE	X		X		X		X		X		X		X		X		X		X
	*Standard Biopsy**	X				X								X						X
	Evaluation	X		X		X		X		X		X		X		X		X		X
Side study	*MRI + targeted biopsies***		*X**			*X*								*X*						*X*
	Evaluation		*X**																	

Figure 3. Prostate cancer Research International: Active surveillance follow-up schedule including MRI.

* MRI 3 months after diagnosis: only targeted biopsies if lesion is visible on MRI (maximum of 3 lesions (2 biopsies per lesion)), no standard TRUS guided biopsies.

** If PSA-doubling time <10 years: An MRI is recommended every year (only in the years no standard biopsy is taken). Additional biopsies are indicated if MRI shows PIRADS progression, more lesions or growth of currently known lesion(s).

Active surveillance endpoints

Clinically relevant endpoints such as (time to) metastasis or disease-specific mortality should preferably be the main outcomes when deciding a patient's treatment strategy [48, 52]. These endpoints imply that a long-term follow-up period of at least a decade is necessary due to the slow-growing nature of PCa [53]. A more practical endpoint is "progression of PCa" as a proxy outcome. However, no uniform definition of disease progression is currently available [53]. Progression can be defined on repeat biopsy findings, or as treatment-free survival. Epstein defines upgrading of the Gleason score at radical prostatectomy as disease progression [54]. Others use biochemically determined recurrence (PSA rise) or presence of distant metastasis indicated by changes in PSA, digital rectal examination (DRE), tumor grade, and tumor volume on biopsy findings, or even MRI [35].

AS is a safe treatment option and should be the primary treatment strategy for low-risk PCa patients with adequate life-expectancy. The risk of metastases during AS is about 4% at 10 years and the risk of death due to PCa is less than 0.5% [55]. Still, after about 5 years, half the men undergoing AS ultimately have curative treatment for their PCa. AS is not suited for all men as some experience anxiety or other psychological side-effects associated with living with untreated cancer.

1.4 PREDICTION MODELS

Follow-up schedules like the ones presented in the previous section call for the development of prediction models that are aimed to determine individual risks of disease outcome. In general, a prediction model is a statistical model used to predict outcomes in the future. Prediction models are used extensively in various fields such as physics [56], meteorology [57], finance [58], technology [59], and medicine [60]. Predictions work because of statistical correlations. A simple correlation is a linear association between two variables, X and Y. For example, the more stairs you climb, the more calories you burn. In daily life, there are many correlations, but these are often difficult to identify as they do not have to be linear. An outcome can be dependent on multiple factors and interactions; therefore, prediction models can come in different shapes and with different underlying statistics. It should be noted that correlation does not equal causality, because causality is more an association and interpretation of the data [61]. In light of this, Simpson's paradox should be taken into consideration when addressing causality. Simpson's paradox is when a correlation appears between X and Y, but disappears or reverses when stratified analyses are performed. This has been addressed by a well-known study which first reported a gender discrimination for acceptance for a university graduation program; 44% of male graduate applicants were accepted in contrast to a 35% acceptance rate of female applicants [62]. However, when examining the individual departments, it appeared that six of the 85 departments were significantly biased against men, whereas only four were significantly biased against women. In conclusion, a confounding variable was found, as women tended to apply to competitive departments with overall low rates of admission, whereas men tended to apply to less competitive departments with high rates of admission.

Prediction models are increasingly used in medicine, especially now we are moving beyond the era of classic evidence-based medicine with groups comparison research to personalized medicine, with a more individualized approach to medical decision-making [63]. Many prediction models, or risk calculators (RC), have been developed for PCa. The ERSPC RC predict PCa on biopsy [64] and had the highest predictive performance in European settings [65]. The RC demonstrated its value by predicting PCa risk and unnecessary test avoidance of up to 30% [64, 66, 67]. Recently, mpMRI findings were included in the ERSPC RC [68]. PSA, DRE, prostate volume, previous biopsy status, age and mpMRI lesion based on Prostate Imaging Reporting and Data System (PIRADS) are used to predict biopsy outcome. A prediction model can take relevant pre-biopsy information into account, but may need to be updated with novel findings or to contemporary center-specific settings to provide accurate estimates on the risk of PCa [69].

The predictive ability of a prediction model is usually evaluated with statistical measures for discrimination and calibration. Discrimination evaluates how well the predicted risks distinguish between patients with and without disease. The c-statistic is the most commonly used measure for discrimination. Calibration evaluates the reliability of the estimated risks: if we predict 10%, on average 10 out of 100 patients should have the disease [70]. However, to answer whether a prediction model improves clinical decision-making, these statistical measures fail to reach a definitive conclusion. Decision curve analysis with net benefit outcomes can identify predictions models that can aid shared-decision-making [71]. In this thesis, the terms net benefit and decision curve analysis will be explained, as well as how they can be used to analyze clinical utility.

Risk-based selection to active surveillance

Several prediction models have been developed for AS to predict PCa progression to improve AS selection. Currently available prediction models and nomograms have a limited predictive ability for progression, with an area-under-the-curve of a ROC diagram never above 0.75, reflecting limited ability to assist in a sharp selection of patients with low-risk PCa for AS [72-79]. Apparently, it is difficult to make a good risk-based selection within the relatively homogeneous groups currently considered for AS. Small groups of men can be found potentially on the borders of the current selection strategies, and may be candidates for WW if very low risk, or for immediate treatment if intermediate risk. Examples of possible candidates for WW may be those with very low PSA values, only 1 core with Gleason 6, and PSAD <0.2. Candidates for treatment might be patients with PSA around 10, 2 cores with Gleason 6, and PSA density (PSA/prostate volume) around 0.2. Furthermore, some models still have to be externally validated prior to considering clinical implication. Further, for patient selection, the risk of the cancer itself needs to be combined with assessments of life expectancy and the anticipated effectiveness of treatment. Hence, a similar risk of progression might lead to treatment in younger men, while it would be acceptable for AS in older men, and for WW in the very old. Moreover, stronger predictors are needed to improve discriminatory performance. Imaging techniques such as MRI are currently under development and novel biomarkers such as PHI, the 4K score, and PCA3 show promising results [80-83]. Ongoing AS cohorts will mature and provide more precise answers in the future. In addition, the Movember Foundation has initiated the Global Action Plan 3 (GAP3). This initiative supports a large centralized database with participating centers from all around the world. This big data initiative aims to create a global consensus on the selection and monitoring of men with low-risk PCa on AS [84]. Optimal selection for AS is complex, and current inclusion criteria vary substantially. Defining low risk either by simple criteria or by more refined risk-based selection models currently has similar results. Future analyses with patients with longer follow-up may allow for more refined inclusion criteria, including new markers.

1.5 AIMS AND OUTLINE OF THE THESIS

Modern medicine should be evidence-based, i.e. effectiveness taken from a group of 'similar' cases, and allow for shared-decision-making for individual patients.
Classic evidence-based medicine is traditionally based on often broad groups of patients qualified for the trial as a reference to provide an outcome for an individual. The average effect of screening in a population, by PSA testing for example, (and the purpose of ERSPC trial) is mostly evaluated by calculating the relative risk of PCa mortality. However, this is different from determining the best screening for an individual. In their state-of-the-art review, Kent, *et al*., explained the fundamental issue and complexity of using group data to guide treatment decisions for individuals [85]. The goal of personalized medicine is to narrow the reference class to produce more patient-centered intervention estimates which support individualized clinical decision-making [85]. A prediction model provides an absolute risk estimate instead of a relative risk, Table 2. For PCa screening, although the ERSPC trial provides strong evidence that screening is effective for at least some men, clinicians still need to understand that the potential screening benefit is influenced by the patient's characteristics and therefore differs between each individual. The overall screening benefit with biomarkers or imaging at population level should be applied on individuals to provide an absolute risk for a specific patient by the use of predictions models. A part of personalized medicine was for example a microsimulation study with the MISCAN simulation study [86]. MISCAN is a decision analytic model that simulates life histories for each individual with and without PSA screening based on the ERSPC data. According to this simulation study, a 4-fold increase in the net benefit of PCa screening can be achieved when older men are screened less and AS is more used for low-risk disease. This is one of the many examples of how to achieve more benefit with personalized medicine.

To translate evidence-based information from the classical group comparison approach to the personalized predictive approach is tried to be accomplished by subgroup analysis. However, a subgroup effect shown for age, say younger/older than 75, does not necessarily mean a different effect for the individual as a patient may still have many other attributes and features of the outcome. Another complication of conventional subgroup analysis is that for statistical convenience the effect is typical tested on a relative scale (e.g. relative risk) rather than an additive scale, i.e. differences between absolute risks [85].

Table 2. Comparison of classic evidence-based medicine with personalized medicine.

	Classic evidence-based medicine	Personalized medicine
Risk estimate Outcome	Relative Odds ratio, relative risk	Absolute Risk percentage of outcome
Study design	Randomized controlled trial or cohort study (small is enough)	Large randomized trial (phase ≥III trials) or large cohort study with aggregation of the overall results using prediction model
Reference class	Wide	Narrow
Decision-making	Population-based decision-making	Patient-specific decision-making

The aim of this thesis is to contribute to the development, evaluation and implementation of personalized evidence-based medicine, with a specific focus on the use of prediction models in the field of PCa. In the first part, the classic evidence-based medicine of PCa screening is described. It starts with the benefit in terms of relative risk reduction of late PCa outcome, and continues with the consequences of negative test results with PSA and prostate biopsy. In addition, the classic evidence-based medicine eligibility criteria of several current active surveillance cohorts and their compliance are assessed.

In the second part of this thesis, I focus on how prediction models can play a role in optimizing the balance of detecting those PCa cases that can benefit from early diagnosis and subsequent treatment versus the reduction of overdiagnosis. Interpretation of the clinical usefulness of prediction models is discussed by explaining the concepts of net benefit and decision curve analysis. Improvements to existing prediction models with the latest International Society of Urological Pathology Gleason grading system, secondary growth patterns like the cribriform growth, and four-kallikrein as a novel biomarker, are studied. In the final section, I discuss implementation issues of prediction models in the daily primary care setting. The general practitioner practice is often the start of the potential PCa screening journey and it is crucial that, at that point, possible survival benefit of early detection and treatment of PCa are correctly assessed. To be of aid in this important phase, a PCa risk calculator is explored that incorporates various risk factors and tumor markers, but also survival probabilities, conditional on comorbidities and treatment benefit.

Research questions addressed in this thesis

Classic evidence-based medicine: Outcomes of prostate cancer screening and active surveillance

- What are the risks of a diagnosis of a clinically significant PCa, metastatic disease and/or PCa death after a false negative screening test or biopsy result in a purely PSA-based screening setting compared to applying an additional test procedure or risk stratification tool? (Chapter 2)

- What is the compliance over time when offering an AS protocol to men with low-risk PCa and how can risk stratification at the start of AS optimize adherence? (Chapter 3)

Towards personalized medicine: How can multivariable prostate cancer prediction models reduce unnecessary testing and support clinical decision-making?

- How do prediction models work and how should the predictions be interpreted in terms of clinical utility? (Chapter 4)

- Can prediction model predicting biopsy outcome be improved by incorporating novel biomarkers and a more refined PCa pathological grading system, and hence decrease the number of unnecessary prostate biopsies and overdiagnosis of potentially indolent disease? (Chapters 5 and 6)

- To what extent can prediction models support triage at primary care practice regarding who receives screening and diagnostic examination, thereby reducing unnecessary testing and overdiagnosis? (Chapter 7)

REFERENCES

1. Wilson AH. The prostate gland: a review of its anatomy, pathology, and treatment. Jama. 2014;312:562.
2. Kumar VL, Majumder PK. Prostate gland: structure, functions and regulation. Int Urol Nephrol. 1995;27:231-43.
3. McNeal JE, Redwine EA, Freiha FS, Stamey TA. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. Am J Surg Pathol. 1988;12:897-906.
4. Egan KB. The Epidemiology of Benign Prostatic Hyperplasia Associated with Lower Urinary Tract Symptoms: Prevalence and Incident Rates. Urol Clin North Am. 2016;43:289-97.
5. Zhou CK, Check DP, Lortet-Tieulent J, Laversanne M, Jemal A, Ferlay J, et al. Prostate cancer incidence in 43 populations worldwide: An analysis of time trends overall and by age group. Int J Cancer. 2016;138:1388-400.
6. Nederlandse kankerregistratie: cijfers over kanker. Accessed through: www.cijfersoverkanker.nl on Feb 2019.
7. Bell KJ, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. Int J Cancer. 2015;137:1749-57.
8. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality, Surveillance Research Program. April 2017.
9. Rao AR, Motiwala HG, Karim OM. The discovery of prostate-specific antigen. BJU Int. 2008;101:5-10.
10. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69:184-210.
11. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol. 2016;40:244-52.
12. Kweldam. CF, Kümmerlin. IP, Nieboer. D, Verhoef EI, Steyerberg. EW, Incrocci. L, et al. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. European Journal of Cancer. 2016.
13. Gandaglia G, Albers P, Abrahamsson PA, Briganti A, Catto JWF, Chapple CR, et al. Structured Population-based Prostate-specific Antigen Screening for Prostate Cancer: The European Association of Urology Position in 2019. European urology. 2019.
14. Cooperberg MR. Implications of the new AUA guidelines on prostate cancer detection in the U.S. Curr Urol Rep. 2014;15:420.
15. Cancer Control: Knowledge into Action: WHO Guide for Effective Programmes: Module 3: Early Detection. Geneva: World Health Organization; 2007.
16. Ciatto S, Zappa M, Villers A, Paez A, Otto SJ, Auvinen A. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. BJU Int. 2003;92 Suppl 2:97-100.
17. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. The Lancet. 2014;384:2027-35.

18. Andriole GL, Crawford ED, Grubb RL, 3rd, Buys SS, Chia D, Church TR, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. Journal of the National Cancer Institute. 2012;104:125-32.

19. Gommers JJJ, Reuvers E, Groenewoud JJM, Broeders MJM, de Vegt F, Verbeek ALM. Changing trends in prostate cancer incidence and mortality rates in the Netherlands since 1990. Abstract book WEON 2019, Annual Netherlands Epidemiology Congress.

20. Etzioni R, Gulati R, Tsodikov A, Wever EM, Penson DF, Heijnsdijk EA, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118:5955-63.

21. Hugosson J, Roobol MJ, Mansson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. European urology. 2019.

22. Bokhorst LP, Bangma CH, van Leenders GJ, Lous JJ, Moss SM, Schröder FH, et al. Prostate-specific antigen-based prostate cancer screening: reduction of prostate cancer mortality after correction for nonattendance and contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. European urology. 2014;65:329-36.

23. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320-8.

24. Pinsky PF, Prorok PC, Yu K, Kramer BS, Black A, Gohagan JK, et al. Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years. Cancer. 2017;123:592-9.

25. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and overtreatment of prostate cancer. European urology. 2014;65:1046-55.

26. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. Jama. 2018;319:1914-31.

27. Bibbins-Domingo K, Grossman DC, Curry SJ. The US Preventive Services Task Force 2017 Draft Recommendation Statement on Screening for Prostate Cancer: An Invitation to Review and Comment. Jama. 2017;317:1949-50.

28. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European urology. 2017;71:618-29.

29. Roobol MJ, Verbeek JFM, van der Kwast T, Kummerlin IP, Kweldam CF, van Leenders G. Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculator for Initial Prostate Biopsy by Incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. European urology. 2017;72:45-51.

30. Tosoian JJ, Carter HB, Lepor A, Loeb S. Active surveillance for prostate cancer: current evidence and contemporary state of practice. Nature reviews Urology. 2016;13:205-15.

31. Bruinsma SM, Bangma CH, Carroll PR, Leapman MS, Rannikko A, Petrides N, et al. Active surveillance for prostate cancer: a narrative review of clinical guidelines. Nature reviews Urology. 2016;13:151-67.

32. Thomsen FB, Brasso K, Klotz LH, Roder MA, Berg KD, Iversen P. Active surveillance for clinically localized prostate cancer--a systematic review. Journal of surgical oncology. 2014;109:830-5.

33. Klotz LH. Active surveillance with selective delayed intervention: walking the line between overtreatment for indolent disease and undertreatment for aggressive disease. Can J Urol. 2005;12 Suppl 1:53-7; discussion 101-2.

34. Thomsen FB, Roder MA, Hvarness H, Iversen P, Brasso K. Active surveillance can reduce overtreatment in patients with low-risk prostate cancer. Danish medical journal. 2013;60:A4575.
35. Loeb S, Bruinsma SM, Nicholson J, Briganti A, Pickles T, Kakehi Y, et al. Active surveillance for prostate cancer: a systematic review of clinicopathologic variables and biomarkers for risk stratification. European urology. 2015;67:619-26.
36. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med. 2016.
37. Lane JA, Donovan JL, Davis M, Walsh E, Dedman D, Down L, et al. Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. The Lancet Oncology. 2014;15:1109-18.
38. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. European urology. 2013;63:597-603.
39. Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol. 2015;33:272-7.
40. Tosoian JJ, Mamawala M, Epstein JI, Landis P, Wolf S, Trock BJ, et al. Intermediate and longer-term outcomes from a prospective active-surveillance program for favorable-risk prostate cancer. J Clin Oncol. 2015;33:3379-85.
41. Welty CJ, Cowan JE, Nguyen H, Shinohara K, Perez N, Greene KL, et al. Extended follow up and risk factors for disease reclassification in a large active surveillance cohort for localized prostate cancer. The Journal of urology. 2015;193:807-11.
42. Eggener SE, Mueller A, Berglund RK, Ayyathurai R, Soloway C, Soloway MS, et al. A multi-institutional evaluation of active surveillance for low risk prostate cancer. The Journal of urology. 2013;189:S19-25; discussion S.
43. Thompson JE, Hayen A, Landau A, Haynes AM, Kalapara A, Ischia J, et al. Medium-term oncological outcomes for extended vs saturation biopsy and transrectal vs transperineal biopsy in active surveillance for prostate cancer. BJU Int. 2015;115:884-91.
44. Godtman RA, Holmberg E, Khatami A, Stranne J, Hugosson J. Outcome following active surveillance of men with screen-detected prostate cancer. Results from the Goteborg randomised population-based prostate cancer screening trial. European urology. 2013;63:101-7.
45. Klotz L. Active surveillance for prostate cancer: patient selection and management. Curr Oncol. 2010;17 Suppl 2:S11-7.
46. Adamy A, Yee DS, Matsushita K, Maschino A, Cronin A, Vickers A, et al. Role of prostate specific antigen and immediate confirmatory biopsy in predicting progression during active surveillance for low risk prostate cancer. The Journal of urology. 2011;185:477-82.
47. Kakehi Y, Kamoto T, Shiraishi T, Ogawa O, Suzukamo Y, Fukuhara S, et al. Prospective evaluation of selection criteria for active surveillance in Japanese patients with stage T1cN0M0 prostate cancer. Jpn J Clin Oncol. 2008;38:122-8.
48. Selvadurai ED, Singhera M, Thomas K, Mohammed K, Woode-Amissah R, Horwich A, et al. Medium-term outcomes of active surveillance for localised prostate cancer. European urology. 2013;64:981-7.
49. Bokhorst LP, Valdagni R, Rannikko A, Kakehi Y, Pickles T, Bangma CH, et al. A Decade of Active Surveillance in the PRIAS Study: An Update and Evaluation of the Criteria Used to Recommend a Switch to Active Treatment. European urology. 2016.
50. Schoots IG, Nieboer D, Giganti F, Moore CM, Bangma CH, Roobol MJ. Is MRI-targeted biopsy a useful addition to systematic confirmatory biopsy in men on active surveillance for low-risk prostate cancer? A systematic review and meta-analysis. BJU Int. 2018.

51. Klotz L. Active surveillance for prostate cancer: overview and update. Curr Treat Options Oncol. 2013;14:97-108.
52. van den Bergh RC, Ahmed HU, Bangma CH, Cooperberg MR, Villers A, Parker CC. Novel tools to improve patient selection and monitoring on active surveillance for low-risk prostate cancer: a systematic review. European urology. 2014;65:1023-31.
53. Welty CJ, Cooperberg MR, Carroll PR. Meaningful end points and outcomes in men on active surveillance for early-stage prostate cancer. Curr Opin Urol. 2014;24:288-92.
54. Epstein JI, Pizov G, Walsh PC. Correlation of pathologic findings with progression after radical retropubic prostatectomy. Cancer. 1993;71:3582-93.
55. Matulewicz RS, Weiner AB, Schaeffer EM. Active Surveillance for Prostate Cancer. JAMA. 2017;318:2152-.
56. Hawking SW. Black hole explosions? Nature. 1974;248:30-1.
57. Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. Nature. 2015;525:47.
58. Ou JA, Penman SH. Financial statement analysis and the prediction of stock returns. Journal of Accounting and Economics. 1989;11:295-329.
59. Baruch J. Steer driverless cars towards full automation. Nature. 2016;536:127.
60. Lloyd-Jones DM, Wilson PWF, Larson MG, Beiser A, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. The American Journal of Cardiology. 2004;94:20-4.
61. Pearl J, Mackenzie D. The Book of Why: The New Science of Cause and Effect. Ingram publisher services us, 2017.
62. Bickel PJ, Hammel EA, O'Connell J W. Sex bias in graduate admissions: data from berkeley. Science. 1975;187:398-404.
63. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Sciences & Business Media, 2019.
64. Roobol MJ, Steyerberg EW, Kranse R, Wolters T, van den Bergh RC, Bangma CH, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. European urology. 2010;57:79-85.
65. Pereira-Azevedo N, Verbeek JFM, Nieboer D, Bangma CH, Roobol MJ. Head-to-head comparison of prostate cancer risk calculators predicting biopsy outcome. Transl Androl Urol. 2018;7:18-26.
66. Roobol MJ, van Vugt HA, Loeb S, Zhu X, Bul M, Bangma CH, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators. European urology. 2012;61:577-83.
67. Roobol MJ, Schröder FH, Hugosson J, Jones JS, Kattan MW, et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group. World J Urol. 2012;30:149-55.
68. Alberts AR, Roobol MJ, Verbeek JFM, Schoots IG, Chiu PK, Osses DF, et al. Prediction of High-grade Prostate Cancer Following Multiparametric Magnetic Resonance Imaging: Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculators. European urology. 2018.
69. Nieboer D, Vergouwe Y, Ankerst DP, Roobol MJ, Steyerberg EW. Improving prediction models with new markers: a comparison of updating strategies. BMC Med Res Methodol. 2016;16:128.
70. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. European urology. 2018;74:796-804.

71. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74.

72. Ankerst DP, Xia J, Thompson IM, Jr., Hoefler J, Newcomb LF, Brooks JD, et al. Precision medicine in active surveillance for prostate cancer: development of the Canary-early detection research Nnetwork active surveillance biopsy risk calculator. European urology. 2015;68:1083-8.

73. Venderbos LD, Roobol MJ, Bangma CH, van den Bergh RC, Bokhorst LP, Nieboer D, et al. Rule-based versus probabilistic selection for active surveillance using three definitions of insignificant prostate cancer. World J Urol. 2016;34:253-60.

74. Wong LM, Neal DE, Finelli A, Davis S, Bonner C, Kapoor J, et al. Evaluation of models predicting insignificant prostate cancer to select men for active surveillance of prostate cancer. Prostate Cancer Prostatic Dis. 2015;18:137-43.

75. Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schröder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. The Journal of urology. 2007;177:107-12; discussion 12.

76. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction of the Pathologic Gleason Score to Inform a Personalized Management Program for Prostate Cancer. European urology. 2016.

77. Truong M, Slezak JA, Lin CP, Iremashvili V, Sado M, Razmaria AA, et al. Development and multi-institutional validation of an upgrading risk tool for Gleason 6 prostate cancer. Cancer. 2013;119:3992-4002.

78. Iremashvili V, Manoharan M, Parekh DJ, Punnen S. Can nomograms improve our ability to select candidates for active surveillance for prostate cancer? Prostate Cancer Prostatic Dis. 2016.

79. Gandaglia G, van den Bergh RCN, Tilki D, Fossati N, Ost P, Surcel CI, et al. How can we expand active surveillance criteria in patients with low- and intermediate-risk prostate cancer without increasing the risk of misclassification? Development of a novel risk calculator. BJU Int. 2018;122:823-30.

80. Park JJ, Park BK. Role of PI-RADSv2 with multiparametric MRI in determining who needs active surveillance or definitive treatment according to PRIAS. J Magn Reson Imaging. 2016.

81. Cantiello F, Russo GI, Cicione A, Ferro M, Cimino S, Favilla V, et al. PHI and PCA3 improve the prognostic performance of PRIAS and Epstein criteria in predicting insignificant prostate cancer in men eligible for active surveillance. World J Urol. 2016;34:485-93.

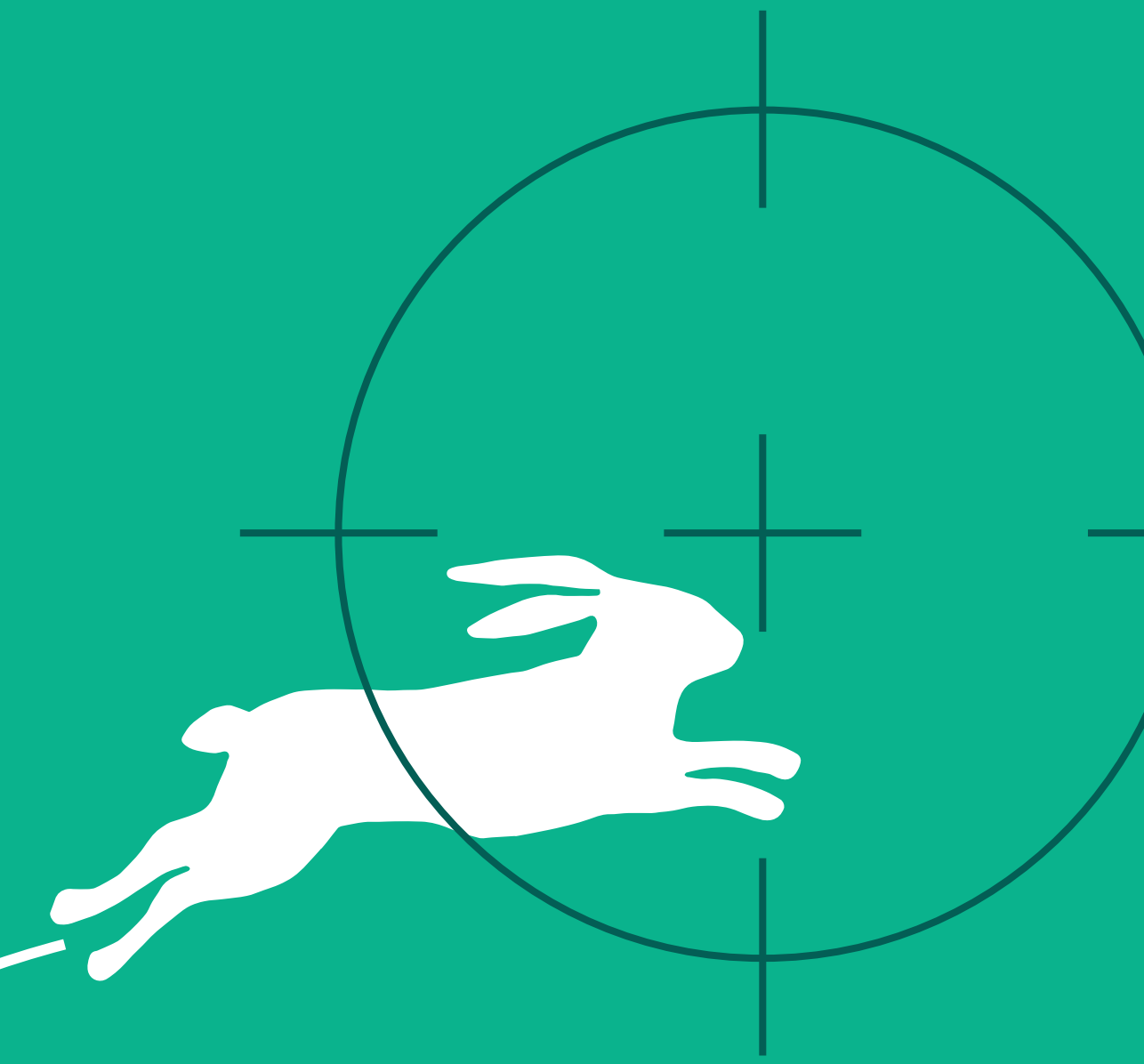
82. Lin DW, Newcomb LF, Brown MD, Sjoberg DD, Dong Y, Brooks JD, et al. Evaluating the four kallikrein panel of the 4kscore for prediction of high-grade prostate cancer in men in the canary prostate active surveillance study. European urology. 2016.

83. Vickers A, Vertosick EA, Sjoberg DD, Roobol MJ, Hamdy F, Neal D, et al. Properties of the four kallikrein panel outside the diagnostic grey zone: meta-analysis of patients with positive digital rectal exam or prostate-specific antigen 10 ng / mL and above. The Journal of urology. 2016.

84. Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW, Nieboer D, et al. The Movember Foundation's GAP3 cohort: a profile of the largest global prostate cancer active surveillance database to date. BJU Int. 2017.

85. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence-based medicine: predictive approaches to heterogeneous treatment effects. Bmj. 2018;363:k4245.

86. Heijnsdijk EA, Wever EM, Auvinen A, Hugosson J, Ciatto S, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med. 2012;367:595-605.



CHAPTER 2

What is an acceptable false negative rate in the detection of prostate cancer?

Verbeek JFM
Roobol MJ
ERSPC Rotterdam Study Group

Transl Androl Urol. 2018 Feb;7(1):54-60

ABSTRACT

Background

In prostate cancer (PCa) screening men and their physicians aim to rule out the presence of potentially life threatening PCa. To date, prostate specific antigen (PSA) testing and systematic prostate biopsy (Bx)-in case of an elevated PSA-are still the main modes of PCa detection. Often uncertainty remains when a PSA-test is <3.0 ng/mL or a Bx shows a benign result, leading to the continuous repeating of procedures. Here we assess the potential consequences of false negatives by studying follow-up data of a purely PSA-based approach with applying sextant Bx, an approach considered to have a high risk of missing PCa diagnosis.

Methods

Our study population consisted of 19,970 men from the ERSPC project section Rotterdam, initially screened in 1993-1999. We assessed clinically significant Gleason ≥3+4 PCa (csPCa) diagnosis within the 4-year screening interval and subsequent screening round 4 years later in men having a PSA <3.0 ng/mL at initial screening (no Bx) and men with Bx (PSA >3.0 ng/mL), but no PCa detected at that time. In addition, we addressed PCa mortality and PCa diagnosis for men with a negative PSA test and negative Bx, who were retested every 4 years covering a 15-year follow-up.

Results

A total of 14,935 men had PSA <3.0 ng/mL in the initial screening round, of whom 75 (0.5%) were diagnosed with csPCa at a subsequent screening examination and 2 (<0.1%) in the 4-year screening interval. For 2,260 men with a previously negative Bx at first screening, the figures were 17 (0.8%) and 2 (0.1%) respectively. Indolent PCa (Gleason ≤3+3) was diagnosed in 312 (2%) men with PSA <3.0 ng/mL initially and 115 (5%) men with initial negative Bx. After a 15-year follow-up, 45 (0.3%) PCa deaths occurred in men with initially low PSA, and 29 men (0.2%) had metastasis. For men with negative Bx, 11 (0.5%) PCa deaths occurred and 4 (0.2%) experienced metastasis.

Conclusions

The false negative rates for men with PSA <3.0 ng/mL and negative sextant Bx are extremely low but not negligible. Proper risk stratification before deciding to biopsy is expected to hardly miss any clinically significant PCa diagnosis. This is especially relevant with the increased use of the relatively expensive multi-parametric magnetic resonance imaging (mpMRI) guided targeted Bx procedures.

INTRODUCTION

In screening practice and case finding systematic transrectal ultrasound biopsies (Bx) are used to detect early prostate cancer (PCa) if prostate specific antigen (PSA) is elevated and/or digital rectal examination (DRE) is abnormal. In the one-size-fits-all approach, and without proper upfront risk stratification, up to 75% of these biopsies turn out to be benign. Hence, these biopsies can be considered unnecessary at that point in time [1]. Diagnostic accuracy of these systematic Bx can be improved by taking more cores [2], and by combining with multi-parametric magnetic resonance imaging (mpMRI) techniques [3, 4]. MpMRI cannot only visualize the difficult to reach PCa lesions located in the anterior and apex region of the prostate, but can also be used as a risk stratification tool before performing a biopsy [5]. As a result, mpMRI is more and more used as the first step in the diagnostic pathway. Although promising, MRI, and if indicated the MRI targeted Bx, is not considered sufficiently accurate to safely replace the systematic approach [6-9], as the negative predictive value of mpMRI varied greatly in a biopsy-naïve group [10]. Thus, currently prostate biopsies consist of at least 12 – 14 cores and are often combined with targeted biopsies. In the diagnostic accuracy discussion, the focus is predominantly on how the number of Bx can be reduced, while little attention is paid to the false negative (FN) aspect of a tool. A FN result means the test is negative for PCa, while in fact the patient has PCa. Prostate cancer, including clinically significant PCa, is not rare among men with low PSA levels [11] or in men with previous negative Bx [12]. Uncertainty remains when the PSA-test result is below the cut-off value or when the Bx shows a benign result, leading to a continuous repeating of procedures which is burdening to the patient and not without risk [13]. Further risk-assessment of asymptomatic men with low PSA avoids unnecessary biopsies, but does not provide a recommendation on how often PSA and DRE should be done [14]. Moreover, no definitive recommendation can be made when to repeat a biopsy if the initial Bx is negative [15]. Additional tools, like PHI, 4Kscore, PCA3, or mpMRI could aid in these uncertain situations, but mainly due to a lack of head-to-head comparisons there are no clear recommendations on which test to use and how to interpret test results. The question remains on what the actual risks are in terms of missing the window of cure when missing or delaying a diagnosis after refraining from biopsy, or having a false negative biopsy result. The use of a purely PSA based algorithm in combination with a sextant biopsy is considered insufficient and at high risk of missing significant PCa diagnoses [16]. To gain insight into the potential benefit from additional tests and repeating biopsy procedures we aimed to assess the (long term) consequences of a PSA test outcome of less than 3.0 ng/ml and negative sextant Bx results in combination with a long retesting interval of 4 year by studying 15-year follow-up data from the European Randomized Study of Screening for Prostate Cancer, section Rotterdam [17].

METHODS

The European Randomized Study of Screening for Prostate Cancer (ERSPC) was established in the 1990s and is the largest randomized study on screening for PCa [18]. In the ERSPC section Rotterdam, a total of 21,210 men were randomized to the screening arm and 19,970 underwent PSA test at the first screening round in 1993–1999. All PSA measurements were performed in a central laboratory with the use of the Beckmann Hybritech assay. During five study rounds, separated by a 4-year screening interval, a PSA level of more than 3.0 ng per milliliter prompted to recommend for a prostate Bx. A false negative PSA test result was defined as a clinically significant diagnosis of Gleason ≥3+4 PCa (csPCa) during the 4-year screening interval or detected at the subsequent screening round in men having a PSA <3.0 ng/mL initially and who did not receive a biopsy at initial screening. A false negative Bx was defined in a similar way: men having had a Bx due to PSA ≥3.0 ng/mL with no PCa detected at the initial round, but with csPCa between the first two rounds or at the second screening examination. Additionally, indolent PCa findings until the subsequent screening round were reported. The transrectal ultrasound (TRUS) sextant biopsy specimens were reviewed as has been described previously [17]. Patients' characteristics between men with false negative PSA vs. men with true negative PSA and false negative Bx vs. true negative Bx results were compared statistically with the chi-square test.

To give an estimation of the clinical impact of the false negative PSA test and Bx we studied the PCa mortality and overall mortality. Mortality rates were derived from patients' survival data available from time of first visit through December 31, 2013. Relevant clinical information for patients who died were presented to a three-blinded committee, whose members had to independently agree on the cause of death; if no agreement was met, the casus was discussed until the cause of death was established or, if not enough information was available, death certificate data was used. The time from first screening visit until PCa death or time to death resulting from other causes was stratified by age; risks of death were computed using cumulative incidence functions with competing risk adjustments for death resulting from PCa and from other causes [19]. Risks of indolent PCa diagnosis, csPCa (Gleason ≥3+4) diagnosis, and progression to metastasis were also computed with cumulative incidence functions. The log-rank test was used for P value calculation to test significance at $P<0.05$. Statistical analyses were performed with R v3.4.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Of the 19,970 men with a PSA test at the initial screening round within the Rotterdam section of the ERSPC, 14,935 (75%) had a PSA <3.0 ng/mL and did not underwent a biopsy. Of them, 75 (0.5%) men were diagnosed with csPCa in a subsequent round (4 years later), and 2 (<0.1%) with csPCa in the 4-year interval between screening rounds. Indolent PCa (Gleason ≤3+3) was diagnosed in 312 (2%) men 4 years after PSA measurement. The total false negative rate of PSA with a cut-off point of 3.0 ng/mL was 2.6% for any PCa, and 0.5% for csPCa.

A total number of 3,249 biopsies were taken in the first screening round due to elevated PSA levels. Negative biopsy results were found in 2,260 (70%) men. In those men, csPCa was found in 17 (0.8%) in the subsequent round 4 years later, and 2 (0.1%) in the 4-year screening interval; 115 (5%) men had indolent PCa. Tables 1 and 2 lists the characteristics of the men with low PSA level and those with negative biopsy stratified to presence of csPCa in the next four years. For men with initially negative BX, age at biopsy was significantly associated with increased risk of PCa diagnosis during the next four years, but family history, DRE and TRUS outcomes were not. There was no association between age at first visit and family history and false negative PSA result. The false negative csPCa were mostly Gleason Score 3+4. The FN rate of Bx was 0.8% for csPCa, and 6% for any PCa.

Table 1. Characteristics of men with PSA <3.0 ng/mL with and without csPCa in a subsequent round.

Characteristic	Men with PSA <3.0 ng/mL without csPCa in subsequent round or interval (n=14,858) (99%)	Men with csPCa on subsequent round or interval (n=77) (1%)	P value
Age at PSA (years), n [%]			
55–59	5,534 [37]	25 [32]	0.39
60–64	4,078 [28]	25 [32]	
65–69	3,199 [22]	20 [26]	
70–74	2,014 [14]	7 [9]	
75+/missing	33 [<1]	0 [0]	
Family history, n [%]			
Positive	971 [7]	8 [10]	0.26
Missing	246 [2]	1 [1]	
Gleason score, n [%]			
≤3+3	312 [2]		
3+4		59 [77]	
4+3		9 [12]	
≥4+4		9 [12]	

PSA = prostate specific antigen; csPCa = clinically significant prostate cancer.

After a 15-year follow-up period (including the possibility of having had three screening visits if still aged <74 years), 45 (0.3%) PCa deaths occurred in men with low initial PSA; 29 men (0.2%) developed metastasis. From these 45 men, 87% had an initial PSA 1–3 ng/mL, whereas 56% of them had a Gleason score 4+4 or higher on diagnostic Bx. Among the 2,260 men with negative Bx, 11 (0.5%) PCa deaths occurred, and 4 (0.2%) experienced metastasis. Five of the 11 man were non-compliant with the screening follow up scheme. Figure 1 illustrates these findings with the competing risks of PCa death and other causes of death as well as PCa diagnosis according to age. The rate of PCa death was not different for the negative PSA test group and negative biopsy group. Age negatively impacted PCa survival and overall survival in men with low initial PSA and negative Bx. From the time of the first visit, PCa incidence increased with a marked increase at time of a screening visit. Finally, it can be inferred from Figure 1 that indolent PCa diagnosis was less in men with a low PSA test compared to men who had a negative biopsy, but that csPCa diagnosis and progression to metastasis were not different.

Table 2. Characteristics of men without presence of PCa on initial biopsy result with csPCa in a subsequent round compared to men without csPCa in a subsequent round.

Characteristic	Men with negative biopsy without csPCa in subsequent round or interval (n=2,241) (99%)	Men with csPCa on subsequent round or interval (n=19) (1%)	P value
Age at biopsy (years), n [%]			
55–59	395 [18]	1 [5]	<0.01
60–64	573 [26]	11 [59]	
65–69	694 [31]	6 [32]	
70–74	567 [25]	1 [5]	
75+/missing	12 [1]	0	
Family history, n [%]			
Positive	155 [7]	1 [5]	0.80
Missing	39 [2]	0 [0]	
DRE, n [%]			0.99
Abnormal	475 [21]	4 [21]	
TRUS, n [%]			0.97
Abnormal	426 [19]	4 [21]	
Gleason score, n [%]			
≤3+3	115 [5]		
3+4		12 [63]	
4+3		1 [5]	
≥4+4		6 [32]	

csPCa = clinically significant prostate cancer.

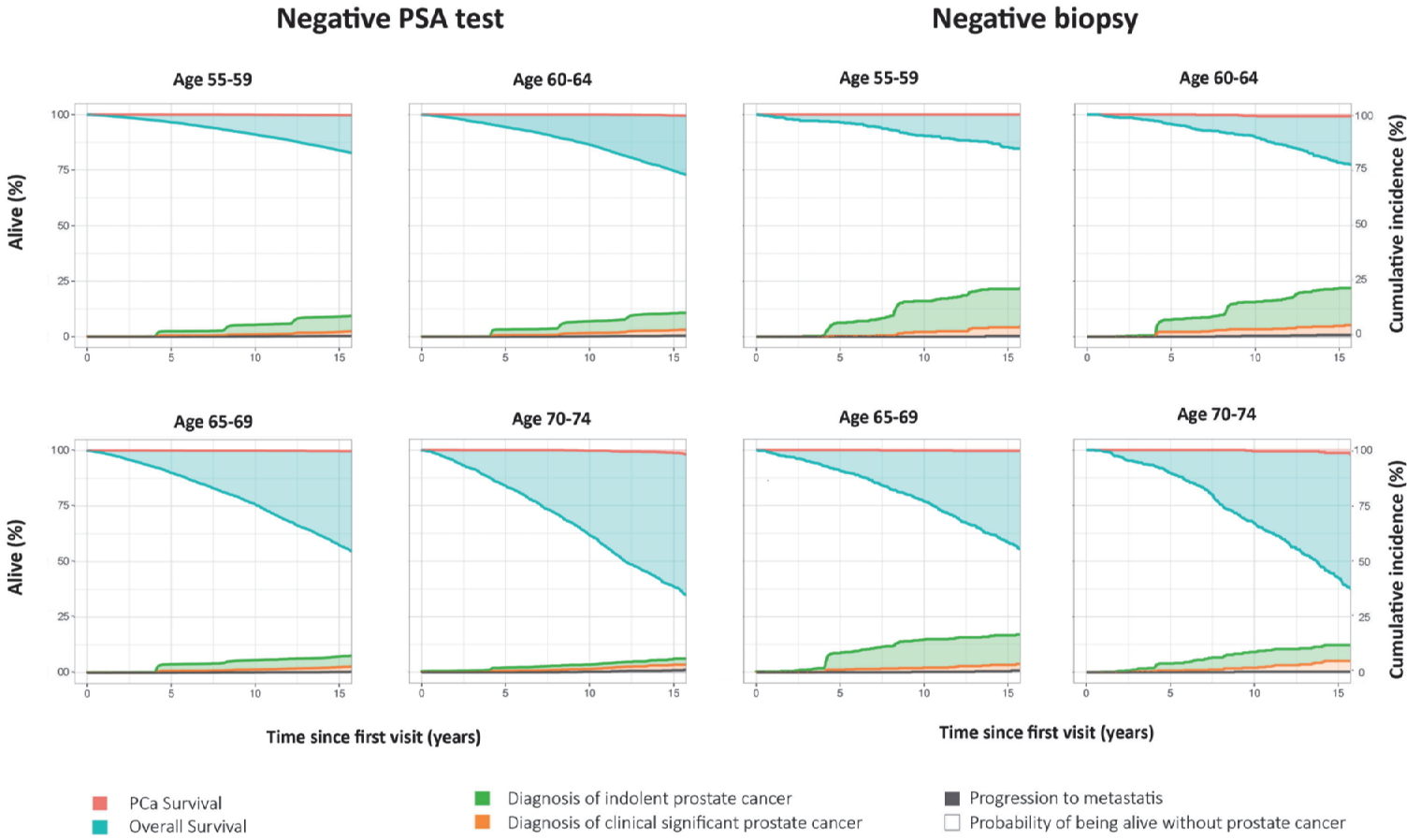


Figure 1. Competing risk of mortality and prostate cancer diagnosis by age in men with a negative PSA (PSA <3.0 ng/ml) or negative prostate biopsy. White area, probability of being alive without prostate cancer; blue area, non–prostate cancer mortality; red area, prostate cancer mortality; green area, cumulative diagnosis of indolent prostate cancer; orange area, cumulative diagnosis of clinical significant prostate cancer (Gleason ≥ 3+4).

DISCUSSION

In today's clinical practice, urologists are anxious to miss a diagnosis of PCa. The decision to perform a random systematic biopsy is based mainly on PSA and DRE results, and the use of risk stratification is not often applied. As a consequence, this approach does not only result in many unnecessary biopsies, but also leaves doubt on the reassurance that PCa is absent or that men are no longer at risk of getting PCa, leading to intensive retesting schemes. Our study, however, showed that the FN rates for PSA <3.0 ng/mL and for sextant Bx are, although not negligible, extremely low.

We showed that PSA screening (including sextant biopsies and applying a long screening interval) detects almost every PCa case that develops within a 15-year period, which means that the maximum achievable increase in detection of potentially life-threatening PCa by applying additional diagnostic tools like novel biomarkers and mpMRI might be limited. Nonetheless, these additional tools should be considered within the broader context of the PSA-screening debate, since a PSA-only screening program in combination with random biopsy sampling results in high rates of unnecessary biopsy and considerable overdiagnosis of indolent PCa [20]. The adoption of proper stratification for high- and low-risk PCa before application of additional diagnostic tools including targeted biopsy, will certainly help in balancing harms and benefits of PCa screening. Moreover, a proper risk stratification, and if indicated adequate imaging and biopsy procedure at the first screening exam, may result in recommendations to refrain from further testing and/or to apply for longer retest intervals if results are benign.

In biopsy naïve men, TRUS Bx directed by mpMRI might improve the detection of PCa [3], however, it is still unclear whether this diagnostic improvement will also lead to a reduction of relevant outcomes like progression to metastasis and PCa mortality. Due to the restricted follow-up period of the available mpMRI study cohorts this cannot yet be evaluated. In our study we used the sextant biopsy procedure, which is known for its poor diagnostic accuracy, anterior lesions for example can easily be missed [2]. Despite this poor accuracy, the PCa mortality at 15 years of follow-up was less than a 0.5%. This is a considerable reduction compared to the national cumulative incidence of PCa death which is 3–5% [21] and the risk of dying on basis of SEER data which show risks of 2.6%, 2.8% and 2.9% for men aged 50, 60 and 70 years, respectively [22]. Note that almost half of the men who died from PCa with a previous negative Bx were not compliant with the follow-up scheme and PCa mortality might be lower with adequate compliance. It should be mentioned that our follow-up results reflect an algorithm with a repeated screening examination every 4 years up to the age of 74. Few csPCa were detected in-between the rounds, which could have been more when a longer screening interval was applied. Furthermore, screening might continue for men age 75 and

over who are in good health, but then individual risk stratification becomes even more crucial due to higher risk of overdiagnosis.

Comparable data on FN rates are available from the Prostate Cancer Prevention Trial (PCPT) having data on PCa prevalence among men with a low PSA level. PCPT is a phase 3, randomized, double-blind, placebo-controlled study designed to determine whether treatment with finasteride could prevent prostate cancer. To study applicability to the general population, only the placebo group of the PCPT was used and they also applied the sextant biopsy method. The reported data show a PCa prevalence for PSA lower than 4 ng/mL of 15% for any PCa, and 2% for clinically significant PCa at 7 years of follow-up, data on mortality are not provided [11]. Our PCa detection rate was lower for indolent and csPCa compared to the results of the PCPT study. This can be explained because we only performed a biopsy when indicated, i.e., for high PSA, and did not perform end-of-study biopsy. In mpMRI studies, comparable data on FN rate in previous negative Bx men are reported [7]. When interpreting these results, it is important to realize that FN rates decrease when PCa prevalence rates increase, and that the prevalence of PCa detection varies with the applied inclusion criteria and biopsy technique [23]. Our study showed that PCa detection increased with follow-up in all age-groups. Therefore, time since the initial benign finding might be of predictive value for when to re-evaluate men with a previous negative Bx and men with low PSA values. This finding, however, could be of limited benefit when the actual evaluation takes place, just as is the case for PSA velocity [24].

In current practice, men with low PSA or previous negative Bx need adequate management to reduce extensive and burdensome testing. This is especially relevant with the increased use of the relatively expensive reflex tests and mpMRI. Instead of improving risk stratification for men with previous negative Bx or men with low PSA, reduction of FN in the first screening moment by a proper risk stratification and improved detection of PCa would reduce anxiety on missing diagnoses considerable and as such lead to a more relaxed follow-up scheme.

CONCLUSION

The false negative rates for men with PSA <3.0 ng/mL and those men with a negative sextant Bx are extremely low, but not negligible. Proper risk stratification before first biopsy in combination with accurate sampling of the prostate if indicated is expected to result in an even further decrease of this FN rate. Perhaps even more important such an approach can reduce the intensity of repeat testing. This is especially relevant with the increased use of relatively expensive reflex tests and mpMRI guided targeted Bx procedures.

REFERENCES

1. Bokhorst LP, Zhu X, Bul M, et al. Positive predictive value of prostate biopsy indicated by prostate-specific-antigen-based prostate cancer screening: trends over time in a European randomized trial*. BJU Int 2012;110:1654-60.
2. Ploussard G, Nicolaiew N, Marchand C, et al. Prospective evaluation of an extended 21-core biopsy scheme as initial prostate cancer diagnostic strategy. Eur Urol 2014;65:154-61.
3. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389:815-22.
4. Alberts AR, Schoots IG, Bokhorst LP, et al. Risk-based patient selection for magnetic resonance imaging-targeted prostate biopsy after negative transrectal ultrasound-guided random biopsy avoids unnecessary magnetic resonance imaging scans. Eur Urol 2016;69:1129-34.
5. Moussa AS, Meshref A, Schoenfield L, et al. Importance of additional "extreme" anterior apical needle biopsies in the initial detection of prostate cancer. Urology 2010;75:1034-9.
6. Haffner J, Lemaitre L, Puech P, et al. Role of magnetic resonance imaging before initial biopsy: comparison of magnetic resonance imaging-targeted and systematic biopsy for significant prostate cancer detection. BJU Int 2011;108:E171-8.
7. Schoots IG, Roobol MJ, Nieboer D, et al. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur Urol 2015;68:438-50.
8. Panebianco V, Barchetti F, Sciarra A, et al. Multiparametric magnetic resonance imaging vs. standard care in men being evaluated for prostate cancer: a randomized study. Urol Oncol 2015;33:17.e1-17.e7.
9. Thompson JE, van Leeuwen PJ, Moses D, et al. The diagnostic performance of multiparametric magnetic resonance imaging to detect significant prostate cancer. J Urol 2016;195:1428-35.
10. Moldovan PC, Van den Broeck T, Sylvester R, et al. What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate Cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel. Eur Urol 2017;72:250-66.
11. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. N Engl J Med 2004;350:2239-46.
12. Valerio M, Anele C, Bott SR, et al. The prevalence of clinically significant prostate cancer according to commonly used histological thresholds in men undergoing template prostate mapping biopsies. J Urol 2016;195:1403-8.
13. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol 2013;64:876-92.
14. Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur Urol 2017;71:618-29.
15. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. J Urol 2013;190:419-26.
16. Scattoni V, Maccagnano C, Capitanio U, Gallina A, Briganti A, Montorsi F. Random biopsy: when, how many and where to take the cores? World J Urol 2014;32:859-69.

17. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med 2009;360:1320-8.
18. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384:2027-35.
19. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Stat Med 2007;26:2389-430.
20. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. Eur Urol 2014;65:1046-55.
21. Central Bureau for Statistics (CBS), Death: underlying cause of death, 2017.www.cbs.nl
22. Surveillance, Epidemiology, and End Results (SEER), Stat Database: Mortality, Surveillance Research Program. April 2017.www.seer.cancer.gov
23. Nelson AW, Harvey RC, Parker RA, et al. Repeat prostate biopsy strategies after initial negative biopsy: meta-regression comparing cancer detection of transperineal, transrectal saturation and MRI guided biopsy. PLoS One 2013;8:e57480.
24. Vickers AJ, Thompson IM, Klein E, et al. A commentary on PSA velocity and doubling time for clinical decisions in prostate cancer. Urology 2014;83:592-6.

CHAPTER 3

Adherence to active surveillance protocols for low-risk prostate cancer: results of the Movember Foundation's Global Action Plan prostate cancer active surveillance (GAP3) initiative

Kalapara AA*
Verbeek JFM*
Nieboer D
Fahey M
Gnanapragasam V
Van Hemelrijck M
Lee LS
Bangma CH
Steyerberg EW
Harkin T
Helleman J
Roobol MJ
Frydenberg M
on behalf of The Movember Foundation's
Global Action Plan Prostate Cancer
Active Surveillance (GAP3) consortium

* These authors are joint first authors

Eur Urol Oncol. 2020 Feb;3(1):80-91
Priority Article

ABSTRACT

Background - Active surveillance (AS) enrolment criteria and follow-up schedules for low-risk prostate cancer vary between institutions. However, uncertainty remains about adherence to these protocols

Objectives - To determine adherence to institution-specific AS inclusion criteria and follow-up schedules within the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) initiative.

Design, settings, and participants - We retrospectively assessed the data of 15 101 patients from 25 established AS cohorts worldwide between 2014 and 2016.

Outcome measurements and statistical analysis - Adherence to individual AS inclusion criteria was rated on a five-point Likert scale ranging from poor to excellent. Nonadherence to follow-up schedules was defined as absence of repeat biopsy 1 yr after the scheduled date. Cohorts were pooled into annual and Prostate Cancer Research International: Active Surveillance (PRIAS)-based biopsy schedules, and a generalized linear mixed model was constructed to test for nonadherence.

Results - Serum prostate-specific antigen (PSA) inclusion criteria were followed in 92%, Gleason score (GS) criteria were followed in 97%, and the number of positive biopsy cores was followed in 94% of men. Both age and tumor stage (T stage) criteria had 99% adherence overall. Pooled nonadherence rates increased over time-8%, 16%, and 34% for annual schedules and 11%, 30%, and 29% for PRIAS-based schedules at 1, 4, and 7 yr, respectively-and did not differ between biopsy schedules. A limitation is that our results do not consider the use of multiparametric magnetic resonance imaging.

Conclusions - In on-going development of evidence-based AS protocols, variable adherence to PSA and GS inclusion criteria should be considered. Repeat biopsy adherence reduces with increased duration of surveillance, independent of biopsy frequency. This emphasizes the importance of risk stratification at the commencement of AS.

Patient summary - We studied adherence to active surveillance protocols for prostate cancer worldwide. We found that inclusion criteria were generally followed well, but adherence to repeat biopsy reduced with time. This should be considered when optimizing future active surveillance protocols.

INTRODUCTION

Since the metamorphosis of active surveillance (AS) from ad hoc to routine, AS has been established as a key management strategy for low-risk localized prostate cancer. Numerous studies have assessed the viability and safety of this approach despite substantial heterogeneity in both inclusion protocols and subsequent monitoring for disease progression [1,2].

As an important facet of prostate cancer management, assessment of the real-world practice of AS and guideline adherence is a key area for research. This has implications on the applicability of the current evidence base supporting the safety and efficacy of AS. Nonadherence to invasive follow-up biopsy regimens is known to be an issue from large national and regional studies [3–5]. However, it remains unclear whether current evidence-based guidelines for AS recruitment and on-going surveillance are congruent with practice.

The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) initiative constitutes the largest worldwide cohort integrating clinical, marker-related, and imaging data of prostate cancer patients on AS. We sought to employ the wealth of data available from centres committed to this consortium to determine adherence to site-specific AS inclusion criteria and follow-up biopsy schedules, and therefore to identify trends that may help establish an optimal surveillance strategy to minimize nonadherence.

PATIENTS AND METHODS

GAP3 has brought together a vast wealth of clinical and research experience in the field of AS for low-risk prostate cancer. Between 2014 and 2016, participating centres uploaded data into a centralized, uniform, and consensus-based AS database via the TranSMART platform [6]. Each participating centre had an active registry of AS patients for 2:2 yr and ethical approval for sharing digital patient data in a centralized global database. In the current analysis, the database contains data of 15 101 patients on AS from 25 centres worldwide, including the USA, Canada, Australasia, the UK, and Europe [7]. All available data from these cohorts were considered in this analysis.

Adherence to AS inclusion criteria

Overall and site-specific clinical and biopsy characteristics of men on AS were summarized using descriptive statistics. In evaluating adherence to AS inclusion criteria, we assessed whether the age, clinical tumor stage (T stage), serum prostate-specific antigen (PSA), diagnostic biopsy Gleason score (GS), PSA density (PSAD), number of positive cores, and maximum extent of biopsy core involvement of included patients met the defined site-specific criteria for selection. We used a five-point Likert scale to rate adherence to institution-specific

inclusion criteria when compared with actual characteristics of included patients: excellent (≥90% of included men met the inclusion criteria), good (80–90%), fair (70–80%), weak (60–70%), or poor (<60%) adherence.

Adherence to follow-up biopsy schedules

We examined adherence to scheduled repeat prostate biopsies. We defined strictly adherent follow-up biopsy as the one occurring within 3 mo before or after the scheduled date per institution-specific protocol. Early biopsies were those performed >3 mo prior to the scheduled date, and late biopsies were those performed >3 mo after. We defined nonadherence as the absence of a scheduled repeat biopsy, where men had sufficient follow-up but did not receive a biopsy within 1 yr after the scheduled date. Our primary outcome was nonadherence.

Cohorts with comparable biopsy schedules were pooled into (1) an annual pool, including those following an annual biopsy protocol, and (2) a Prostate Cancer Research International: Active Surveillance (PRIAS)-based pool, including cohorts that followed a PRIAS or a similar protocol (repeat biopsy at 1, 4, and 7 yr). We constructed a generalized linear mixed model with nonadherence versus adherence (including strictly adherent, early, or late biopsy) outcomes. We used random effect for patient and centre, and fixed effect for time on AS and follow-up schedule (annual or PRIAS based), to test for nonadherence during follow-up and differences between AS schedules.

The database was frozen for analysis in November 2017 (version 2.3.1). All statistical analyses were performed using R version 3.4.2 (R Foundation, Vienna, Austria).

RESULTS

Adherence was determined relative to site-specific inclusion criteria and re-biopsy schedules (Tables 1 and 2) [8]. In total, 15 101 patients were included for analysis. Clinical and biopsy features of these patients are listed in Table 3, stratified by centre.

Adherence to AS inclusion criteria

Adherence rates to site-specific inclusion criteria, both overall and by centre, are shown in Fig. 1. Six centres were excluded from analysis of overall adherence, due to missing clinical T-stage data in three and extent of core positivity in three. Of the remaining 19 centres, overall adherence to all site-specific inclusion criteria was excellent (≥90%) for seven centres, good (80–90%) for four centres, fair (70–80%) for four centres, and weak (60–70%) for four centres. Poor adherence (<60%) was not observed. All eight sites with fair or weak overall adherence had fair-weak adherence in no more than one individual criterion.

Considering each parameter separately, adherence to PSA was excellent in 14 centres, good in seven centres, fair in one centre, and weak in one centre. In total, PSA inclusion criteria were followed in 11 956 of 13 010 (92%) men with known serum PSA level.

For GS criteria, adherence across institutions was excellent in 21 and good in four centres. Overall, GS adherence was observed in 14 314 of 14 808 (97%) included men. All 12 institutions that allowed GS 3 + 4 = 7 at diagnosis had excellent adherence. Eight of these 12 institutions considered the number of positive biopsy cores, of which six had excellent, one had weak, and one had poor adherence. Six of these 12 institutions considered the extent of core positivity, of which three had excellent, one had good, and one had fair adherence, and one had missing data. Conversely, the four centres with good adherence permitted only GS 6 and PSA < 10 at inclusion, with the extent of core positivity not considered in three of four centres. Adherence to the criteria related to age (overall 12 129 [99%] of 12 181 men) and clinical T stage (overall 10 721 [99%] of 10 830 men) was excellent for all centres except for one, which showed good adherence. Fifteen centres considered the number of positive cores in their inclusion criteria, with adherence in 8102 of 8634 (94%) men, and 11 centres considered the maximum extent of core positivity, with adherence in 6686 of 7190 (93%) men.

Table 1 – Active surveillance inclusion criteria by centre in GAP3.

Centre	Age (yr)	Clinical stage	Serum PSA (ng/ ml)	Biopsy Gleason score	Serum PSA density (ng/ml/g)	Positive cores (*n*)	Min-max extent cancer per core
Asia/Australia							
MEASCAP—Melbourne	>18	T1c or T2	≤10	3+3=6 or 3+4=7, up to 20% of pattern 4	<0.2	1–2	≤5 mm per core (±equal 25% of a core)
KU—Kagawa	50–80	T1c N0M0	≤20	≤3+3=6	NR	1–2 per 6–12 systematic biopsy cores	≤50%
SGH—Singapore	>18	T1-T2	<10	3+3	NR	NR	≤50%
YUHS—Seoul	>18	T1-T2a	≤10	≤7	NR	<2	NR
Europe							
HUCH—Helsinki	NR	≤T2	≤10	≤6	<0.2	1–2	NR
SU—Gothenburg	>18	T1	<10	3+3=6	NR	NR	NR
SUS—Malmö	>18	T1c or T2	≤10	3+3=6 or 3+4	<0.2	1–2	NR
UCD— Dublin	>18	T1-T2a	<10	≤6	NR	NR	NR
EMC, Rotterdam	NR	≤T2	≤10	≤6	<0.2	1–2	NR
CHU— Lille	>18	T1c	≤15	3+3=6 or 3+4	NR	≤3	1–5
KB— Baden	>18	T1a-T1c (T2a)	<10	≤6 (3+3)	NR	≤2	≤5 mm/core
INT—Milan	>18	PRIAS: T1c-T2a; SAINT: T1c-T2a; T2b if ≤0.5 ml tumour volume and negative peripheral zone biopsy	≤10	3+3=6	PRIAS: <0.20; SAINT: NR	PRIAS: ≤2; ≤15% if saturation biopsy; No restriction if GS 3+3 fusion biopsy or negative RMmp; SAINT: ≤3 and ≤25% of total cores	PRIAS: ×; SAINT: ≤50%
IVO— Valencia	<80	T1a, T1b, T1c	≤10	≤6 or 3+4 (3+4 for men >70 yr old)	<0.2	≤2	33–≤50%
UK							
CUHT—Cambridge	50–75	T1-T2a	≤10	≤6; 7 based on patient-clinical discussion	NR	NR	NR
GSTT—London	>18	T1a-T1b, T2	≤15	≤6 or ≤3+4=7	NR	NR	NR
UCL—London	>18	NR	<20	Up to Gleason 7	No specific limit	Targeted biopsy strategy used	NR
USA							
MSKCC—New York, NY	>18	NR	NR	6	NR	NR	≤50%
JHU—Baltimore, MD	>40	T1c, T2a	<10 (for men not meeting VLR criteria)	3+3=6	<0.15 to define VLR, and <0.1 if PSA over 10 ng/ml	≤2	≤50% of any core unless unilateral disease, then NR
MUSIC—Michigan	>18	T1-T2b	≤10	<7	NR	≤1/3 or all cores involved	≤50%
EU—Atlanta, GA	>18	T1-T2	<10	1. 3+3=6	<0.15	1. <6 if 3+3=6	<50% any core
				2. 3+4=7 with <10% of pattern 4 if age >70		2. <3 if 3+4=7	
MDACC—Houston, TX	>18	T1-T2	<4	3+3=6	NR	≤1 with <3 mm tumour	NR
			<4	3+4=7		≤1 with <2 mm tumour	
UCSF—San Francisco, CA	>18	≤T2	≤10	≤6	NR	NR	NR
Canada							

Table 1 (*Continued*)

Centre	Age (yr)	Clinical stage	Serum PSA (ng/ ml)	Biopsy Gleason score	Serum PSA density (ng/ml/g)	Positive cores (*n*)	Min-max extent cancer per core
UOFC—Calgary	<59 (stage I)	≤T2	NR	6 (stage I or II): or 3 + 4 stage II	NR	≤3 stage I	<50%
	>60 (stage II)					<6 stage II	
UOFT—Toronto	>18	NR	≤10	≤6	NR	NR	<30%
			10–20	3 + 4			
UBC—Vancouver	NR	≤T2	≤10	≤6	<0.2	1- 2	Maximum of 50% or 5 mm of PCa in a single core

CHU = Lille University Hospital Center, Lille, France; CUHT = Cambridge University Hospitals NHS Trust, Cambridge, UK; EMC = Erasmus Medical Center, Rotterdam, the Netherlands; EU = Emory University School of Medicine, Winship Cancer Institute, Atlanta, GA, USA; GAP3 = Global Action Plan Prostate Cancer Active Surveillance; GS = Gleason score; GSTT = Guy's and St Thomas' NHS Foundation Trust, London, UK; HUCH = Helsinki University Central Hospital, Helsinki, Finland; INT = Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; IVO = Instituto Valenciano de Oncología, Valencia, Spain; JHU = Johns Hopkins University, Baltimore, MD, USA; KB = Kantonsspital Baden, Baden, Switzerland; KU = Kagawa University Faculty of Medicine, Kagawa, Japan; MDACC = MD Anderson Cancer Centre, Houston, TX, USA; MEASCAP = Monash University and Epworth Health, Melbourne, Australia; MSKCC = Memorial Sloan Kettering Cancer Center, New York, NY, USA; MUSIC = University of Michigan and Michigan Urological Surgery Improvement Collaborative, Michigan, USA; NR = not reported; PCa = prostate cancer; PRIAS = Prostate Cancer Research International: Active Surveillance; PSA = prostate-specific antigen; SAINT = Sorveglianza Attiva Istituto Nazionale Tumori; SGH = Singapore General Hospital, Singapore; SU = Sahlgrenska University Hospital, Göteborg, Sweden; SUS = Skåne University Hospital, Malmö, Sweden; UBC = University of British Columbia, BC Cancer Agency, Vancouver, Canada; UCD = University College Dublin, Dublin, Ireland; UCL = University College London & University College London Hospitals Trust, London, UK; UCSF = University of California, San Francisco, CA, USA; UOFC = University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada; UOFT = University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada; VLR = very low risk; YUHS = Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea.

Table 2 – Active surveillance follow-up protocol by centre in GAP3.

Centre	Serum PSA	PSA kinetics (PSADT/PSAV)	DRE	Biopsy	mpMRI
Asia/Australia					
MEASCAP—Melbourne	Every 3 mo	NR	Every 6 mo	After months 12, 48, and 84	NR
KU—Kagawa	Every 3 mo for the first 6 mo, then every 3 mo	NR	Every 12 mo	Every 12 mo	NR
SGH—Singapore	3–6 monthly for the first 2 yr, 6–12 monthly thereafter	Every 12 mo	Every 12 mo	Every 12 mo	Every 12 mo
YUHS—Seoul	Every 3 mo	NR	NR	Considered if mpMRI result is changed	Every 12 mo
Europe					
HUCH—Helsinki	Every 3 mo	Every 6 mo	Every 6 mo	After months 12, 48, and 84	NR
SU—Gothenburg	Every 3–6 mo	NR	Every 6–12 mo	Every 2–3 yr	NR
SUS—Malmö	Every 3 mo	NR	Every 6 mo	After months 12, 48, and 84	NR
UCD—Dublin	NR	NR	NR	1 yr, then every other year	NR
EMC—Rotterdam	Every 3 mo	Every 6 mo	Every 6 mo	After months 12, 48, and 84	NR
CHU—Lille	Every 6 mo	NR	Every 12 mo	At month 12	At month 12
KB—Baden	Every 6 mo	NR	Every 6 mo	Every 24 mo	NR
INT—Milan	Every 3 mo	NR	Every 6 mo	Every 12 mo for the first 2 yr and then every 24 mo	NR
IVO—Valencia	Every 6 mo	NR	Every 6 mo	Month 24 from start on AS, then every 3 yr if no progression	NR
UK					
CUHT—Cambridge	Every 3 mo	Every 12 mo	NR	At months 12, 36, and 60	At months 12, 36, and 60
GSTT—London	Every 6 mo	Every 12 mo	Every 12 mo	NR	Every 12 mo
UCL—London	3–4 monthly in 1 yr; 6 monthly after that	NR	Not routinely done	For men where there is a change in MRI and uncertainty about converting to active treatment	At baseline and 12 mo for all men. After that, dependent on risk factors including MRI findings, PSA density, and Gleason score
USA					
MSKCC—New York, NY	Every 6 mo	NR	Every 6 mo	Every 3 yr	Every 18 mo
JHU—Baltimore, MD	Every 6 mo	NR	Every 6 mo	Every 12 mo	NR
MUSIC—Michigan,	Every 3–6 mo	NR	Every 12 mo	Every other year	Every other year; confirmatory test in first 3–4 mo
EU—Atlanta, GA	Every 6 mo	NR	Every 12 mo	Every 12 mo	Annually for the first 3 yr, then final scheduled at 5 yr
MDACC—Houston, TX	Every 6 mo	NR	Every 6 mo	Every 12 mo	Every 12 mo
UCSF—San Francisco, CA	Every 3 mo	NR	Every 6 mo	Every 12–24 mo	NR
Canada					
UBC—Vancouver	Every 3 mo	Every 6 mo	Every 6 mo	After months 12, 48, and 84	NR
UOFC—Calgary	Every 6 mo	NR	Every 6 mo	At year 1, then every 2 yr	When PSA > 10
UOFT—Toronto	Every 3 mo until 2 yr, then every 6 mo	Every 12 mo	Every 6 mo	At years 1, 4, 7, 10, and 15	Every 12 mo

AS = active surveillance; CHU = Lille University Hospital Center, Lille, France; CUHT = Cambridge University Hospitals NHS Trust, Cambridge, UK; DRE = digital rectal examination; EMC = Erasmus Medical Center, Rotterdam, the Netherlands; EU = Emory University School of Medicine, Winship Cancer Institute, Atlanta, GA, USA; GAP3 = Global Action Plan Prostate Cancer Active Surveillance; GSTT = Guy's and St Thomas' NHS Foundation Trust, London, UK; HUCH = Helsinki University Central Hospital, Helsinki, Finland; INT = Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; IVO = Instituto Valenciano de Oncología, Valencia, Spain; JHU = Johns Hopkins University, Baltimore, MD, USA; KB = Kantonsspital Baden, Baden, Switzerland; KU = Kagawa University Faculty of Medicine, Kagawa, Japan; MDACC = MD Anderson Cancer Centre, Houston, TX, USA; MEASCAP = Monash University and Epworth Health, Melbourne, Australia; mpMRI = multiparametric magnetic resonance imaging; MRI = magnetic resonance imaging; MSKCC = Memorial Sloan Kettering Cancer Center, New York, NY, USA; MUSIC = University of Michigan and Michigan Urological Surgery Improvement Collaborative, Michigan, USA; NR = not reported; PSA = prostate-specific antigen; PSADT = PSA doubling time; PSAV = PSA velocity; SGH = Singapore General Hospital, Singapore; SU = Sahlgrenska University Hospital, Göteborg, Sweden; SUS = Skåne University Hospital, Malmö, Sweden; UBC = University of British Columbia, BC Cancer Agency, Vancouver, Canada; UCD = University College Dublin, Dublin, Ireland; UCL = University College London & University College London Hospitals Trust, London, UK; UCSF = University of California, San Francisco, CA, USA; UOFC = University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada; UOFT = University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada; YUHS = Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea.

Table 3 – Clinical and biopsy characteristics of patients in GAP3, stratified by site.

	N	Age (yr) Median (IQR)	Clinical stage, T2 or higher *N* (%)	Serum PSA (ng/ml) Median (IQR)	Biopsy Gleason score *N* (%)		Prostate volume (ml) Median (IQR)	Positive cores (*n*) Median (IQR)	Max % of cancer in any one core Median (IQR)
EU—Atlanta	61	67 (59, 71)	4 (11)	5.40 (4.04, 7.21)	3+3	49 (83)	44.0 (30.5, 61.0)	1 (1, 2)	10.0 (5.0, 29.3)
					3+4	6 (10)			
					4+3	2 (3)			
					≥4+4	2 (3)			
					Missing	2			
KB—Baden	193	66 (61, 70)	11 (6)	4.57 (2.93, 6.80)	3+3	189 (98)	37.7 (29.7, 50.2)	NA	NA
					3+4	2 (1)			
					4+3	1 (1)			
					≥4+4	0			
					Missing	1			
UOFC—Calgary	581	62 (57, 67)	63 (11)	4.27 (2.70, 6.40)	3+3	541 (94)	NA	1 (1, 2)	10.0 (5.0, 20.0)
					3+4	33 (6)			
					4+3	3 (0)			
					≥4+4	0			
					Missing	5			
CUHT—Cambridge	291	67 (61, 72)	55 (11)	7.23 (5.60, 10.01)	3+3	241 (83)	50.0 (37.0, 66.0)	2 (1, 3)	5.0 (2.0, 10.0)
					3+4	43 (15)			
					4+3	6 (2)			
					≥4+4	1 (0.3)			
					Missing	0			
UCD—Dublin	53	66 (57, 71)	NA	5.56 (4.32, 6.70)	3+3	38 (81)	NA	1 (1, 2)	5.0 (5.0, 10.0)
					3+4	7 (15)			
					4+3	1 (2)			
					≥4+4	1 (2)			
					Missing	6			
EMC—Rotterdam	2422	66 (61, 71)	324 (13)	5.70 (4.60, 7.20)	3+3	2412 (99.7)	45.0 (35.0, 60.0)	1 (1, 2)	NA
					3+4	8 (0.3)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	2			
SU—Gothenburg	1001	68 (65, 71)	183 (20)	4.90 (3.65, 7.50)	3+3	799 (82)	37.2 (29.0, 50.9)	1 (1, 2)	NA
					3+4	138 (14)			
					4+3	32 (3)			
					≥4+4	11 (1)			
					Missing	21			
HUCH—Helsinki	303	63 (60, 68)	1 (0)	5.60 (4.40, 6.93)	3+3	289 (100)	40.4 (33.2, 50.0)	1 (1, 2)	NA
					3+4	0 (0)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	14			
JHU—Baltimore, MD	1457	66 (62, 69)	NA	4.80 (3.60, 6.27)	3+3	1428 (99.6)	45.0 (35.0, 60.0)	1 (1, 2)	5.0 (1.0, 10.0)
					3+4	6 (0.4)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	23			
KU—Kagawa	118	70 (66, 73)	0 (0)	6.50 (5.10, 8.97)	3+3	118 (100)	37.6 (26.6, 48.8)	1 (1, 1)	11.2 (6.9, 18.4)
					3+4	0 (0)			

Table 3 (*Continued*)

	N	Age (yr) Median (IQR)	Clinical stage, T2 or higher *N* (%)	Serum PSA (ng/ml) Median (IQR)	Biopsy Gleason score *N* (%)		Prostate volume (ml) Median (IQR)	Positive cores (*n*) Median (IQR)	Max % of cancer in any one core Median (IQR)
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	0			
CHU—Lille	166	65 (60, 71)	22 (13)	6.59 (5.00, 8.30)	3+3	163 (98)	57.0 (40.0, 78.5)	1 (1, 2)	10.0 (8.3, 17.8)
					3+4	3 (2)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	0			
GSTT—London	399	63 (58, 68)	145 (36)	6.00 (4.00, 9.00)	3+3	205 (70)	42.5 (30.0, 59.8)	1 (1, 3)	20.0 (10.0, 40.0)
					3+4	69 (24)			
					4+3	15 (5)			
					≥4+4	3 (1)			
					Missing	107			
UCL—London	390	62 (57, 67)	NA	6.30 (4.55, 8.89)	3+3	306 (84)	49.0 (35.0, 68.3)	2 (2, 3)	15.0 (10.0, 30.0)
					3+4	57 (15)			
					4+3	2 (0.6)			
					≥4+4	1 (0.3)			
					Missing	24			
SUS—Malmö	137	66 (63, 69)	25 (18)	5.00 (4.00, 6.30)	3+3	126 (92)	41.8 (33.5, 57.0)	1 (1, 2)	12.1 (5.8, 18.7)
					3+4	10 (7.3)			
					4+3	1 (0.7)			
					≥4+4	0 (0)			
					Missing	0			
MDACC—Houston, TX	185	63 (57, 70)	22 (12)	3.20 (2.10, 4.30)	3+3	178 (99.4)	43.8 (29.7, 55.3)	1 (1, 1)	NA
					3+4	1 (0.6)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	6			
MEASCAP—Melbourne	283	64 (58, 68)	39 (15)	5.90 (4.50, 8.35)	3+3	283 (100)	49.5 (38.3, 66.0)	2 (1, 3)	NA
					3+4	0 (0)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	0			
INT—Milan	722	66 (61, 71)	58 (8)	5.70 (4.50, 7.20)	3+3	715 (99)	46.0 (35.0, 63.0)	1 (1, 2)	10.0 (5.0, 20.0)
					3+4	7 (1)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	0			
MUSIC—Michigan	1973	65 (60, 71)	288 (15)	5.32 (4.17, 7.18)	3+3	1601 (81)	43.7 (32.0, 60.9)	2 (1, 3)	10.0 (5.0, 25.0)
					3+4	296 (15)			
					4+3	56 (3)			
					≥4+4	18 (1)			
					Missing	2			
MSKCC—New York, NY	1085	64 (58, 68)	154 (15)	4.63 (3.40, 6.26)	3+3	1075 (99.1)	48.4 (36.3, 63.5)	1 (1, 2)	5.0 (3.0, 12.0)
					3+4	9 (0.8)			
					4+3	1 (0.1)			
					≥4+4	0 (0)			
		0	40	177	Missing	0			

Table 3 (*Continued*)

	N	Age (yr) Median (IQR)	Clinical stage, T2 or higher *N* (%)	Serum PSA (ng/ml) Median (IQR)	Biopsy Gleason score *N* (%)		Prostate volume (ml) Median (IQR)	Positive cores (*n*) Median (IQR)	Max % of cancer in any one core Median (IQR)
YUHS—Seoul	60	70 (63, 74)	NA	4.42 (2.82, 6.46)	3+3	51 (85)	43.2 (31.7, 56.8)	3 (2, 7)	10.0 (5.0, 22.0)
					3+4	7 (11.7)			
					4+3	2 (3.3)			
					≥4+4	0 (0)			
					Missing	0			
SGH—Singapore	250	67 (62, 72)	18 (7)	6.20 (4.26, 8.60)	3+3	213 (85.2)	31.0 (23.0, 43.5)	1 (1, 2)	15.0 (5.0, 30.0)
					3+4	33 (13.2)			
					4+3	4 (1.6)			
					≥4+4	0 (0)			
					Missing	0			
UOFT—Toronto	1051	67 (61, 72)	165 (17)	5.78 (4.03, 7.79)	3+3	917 (88.6)	NA	2 (1, 3)	15.0 (5.0, 30.0)
		0			3+4	118 (11.4)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	20			
UCSF—San Francisco, CA	1500	63 (57, 68)	450 (30)	5.44 (4.20, 7.40)	3+3	1311 (88.4)	37.0 (28.0, 52.0)	2 (1, 3)	17.0 (10.0, 33.0)
					3+4	144 (9.7)			
					4+3	24 (1.6)			
					≥4+4	4 (0.3)			
					Missing	17			
IVO—Valencia	351	67 (61, 71)	30 (9)	5.55 (4.00, 7.80)	3+3	327 (93.2)	40.0 (31.0, 51.3)	1 (1, 2)	NA
					3+4	17 (4.8)			
					4+3	7 (2)			
					≥4+4	0 (0)			
					Missing	0			
UBC—Vancouver	69	67 (60, 71)	15 (22)	5.32 (3.00, 7.10)	3+3	69 (100)	43.4 (30.0, 58.0)	1 (1, 1)	NA
					3+4	0 (0)			
					4+3	0 (0)			
					≥4+4	0 (0)			
					Missing	0			
Total	15 101	65 (60, 70)	2046 (16)	5.40 (4.02, 7.27)	3+3	13 598 (92)	43.2 (33.0, 59.0)	1 (1, 2)	10.0 (5.0, 20.0)
					3+4	1015 (7)			
					4+3	156 (1)			
					≥4+4	40 (0)			
					Missing	292			

CHU = Lille University Hospital Center, Lille, France; CUHT = Cambridge University Hospitals NHS Trust, Cambridge, UK; EMC = Erasmus Medical Center, Rotterdam, the Netherlands; EU = Emory University School of Medicine, Winship Cancer Institute, Atlanta, GA, USA; GAP3 = Global Action Plan Prostate Cancer Active Surveillance; GSTT = Guy's and St Thomas' NHS Foundation Trust, London, UK; HUCH = Helsinki University Central Hospital, Helsinki, Finland; INT = Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; IQR = interquartile range; IVO = Instituto Valenciano de Oncología, Valencia, Spain; JHU = Johns Hopkins University, Baltimore, MD, USA; KB = Kantonsspital Baden, Baden, Switzerland; KU = Kagawa University Faculty of Medicine, Kagawa, Japan; MDACC = MD Anderson Cancer Centre, Houston, TX, USA; MEASCAP = Monash University and Epworth Health, Melbourne, Australia; MSKCC = Memorial Sloan Kettering Cancer Center, New York, NY, USA; MUSIC = University of Michigan and Michigan Urological Surgery Improvement Collaborative, Michigan, USA; NA = not available; SGH = Singapore General Hospital, Singapore; PSA = prostate-specific antigen; SU = Sahlgrenska University Hospital, Göteborg, Sweden; SUS = Skåne University Hospital, Malmö, Sweden; UBC = University of British Columbia, BC Cancer Agency, Vancouver, Canada; UCD = University College Dublin, Dublin, Ireland; UCL = University College London & University College London Hospitals Trust, London, UK; UCSF = University of California, San Francisco, CA, USA; UOFC = University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada; UOFT = University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada; YUHS = Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea.

Cohort - n	Adherence %							
	Age	Clinical T-stage	PSA	Biopsy Gleason Score (GS)	PSAD	No. positive biopsies	Min-Max extent cancer per core	Overall Adherence
United States								
JHU-Baltimore - 1457	100		94	100	79	94	96	
UCSF-San Francisco - 1500	100	100	89	88				80
MSKCC-New York - 1085	100			99			99	99
MDACC-Houston - 185	100	100	69	100		100		69
EU-Atlanta - 61	100	100	95	93	63	98	88	63
MUSIC-Michigan - 1973	100	98	91	99		86	92	84
Canada								
UOFC-Calgary - 581	100	100		100		97	96	96
UOFT-Toronto - 1051	100		88	100			74	74
UBC-Vancouver - 69		100	99	100	96	97		
Europe								
KB-Baden - 193	100	100	91	98				
UCD-Dublin - 53	100		92	81				
EMC-Rotterdam - 2422		100	100	100	100	100		99
SU-Gothenburg - 1001	100	99	84	82				70
HUCH-Helsinki - 303		100	100	100	100	100		100
INT-Milan - 722	100	100	97	99	70	98	97	70
IVO-Valencia - 351	100	91	87	97	75	92		
CHU-Lille - 166	100	87	95	100		97	99	82
CUHT-Cambridge - 291	82	98	75	98				64
SUS-Malmö - 137	100	100	99	99	100	100		99
GSTT-London - 399	100	100	92	94				88
UCL-London - 390	100		100	100				99
Asia/Australia								
MEASCAP-Melbourne - 283	100	100	87	100	91	68	100	68
SGH-Singapore - 250	100	100	84	85			92	72
YUHS-Seoul - 60	100		89	100		37		
KU-Kagawa - 118	100	100	100	100		100	100	100
Absolute adherence - n (%)	12129 (100)	10721 (99)	11956 (92)	14314 (97)	5184 (87)	8102 (94)	6686 (93)	9021 (87)
Total included - n	12181	10830	13010	14808	5944	8634	7190	10354*
No criteria - n (%)	2794 (19)	2526 (17)	1666 (11)	0 (0)	8295 (59)	5630 (40)	6741 (48)	NA
Missings - n (%)	126 (1)	1745 (12)	427 (3)	299 (2)	862 (2)	837 (6)	1170 (8)	5466

** including cohorts without criteria*

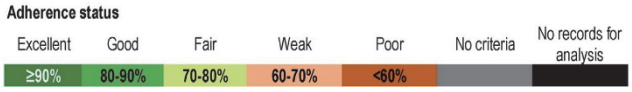


Figure 1 – Adherence rates to site-specific inclusion criteria, both overall and by centre.
CHU = Lille University Hospital Center, Lille, France; CUHT = Cambridge University Hospitals NHS Trust, Cambridge, UK; EMC = Erasmus Medical Center, Rotterdam, the Netherlands; EU = Emory University School of Medicine, Winship Cancer Institute, Atlanta, GA, USA; GSTT = Guy's and St Thomas' NHS Foundation Trust, London, UK; HUCH = Helsinki University Central Hospital, Helsinki, Finland; INT = Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; IVO = Instituto Valenciano de Oncología, Valencia, Spain; JHU = Johns Hopkins University, Baltimore, MD, USA; KB = Kantonsspital Baden, Baden, Switzerland; KU = Kagawa University Faculty of Medicine, Kagawa, Japan; MDACC = MD Anderson Cancer Centre, Houston, TX, USA; MEASCAP = Monash University and Epworth Health, Melbourne, Australia; MSKCC = Memorial Sloan Kettering Cancer Center, New York, NY, USA; MUSIC = University of Michigan and Michigan Urological Surgery Improvement Collaborative, Michigan, USA; NA = not available; PSA = prostate-specific antigen; PSAD = prostate-specific antigen density; SGH = Singapore General Hospital, Singapore; SU = Sahlgrenska University Hospital, Göteborg, Sweden; SUS = Skåne University Hospital, Malmö, Sweden; UBC = University of British Columbia, BC Cancer Agency, Vancouver, Canada; UCD = University College Dublin, Dublin, Ireland; UCL = University College London & University College London Hospitals Trust, London, UK; UCSF = University of California, San Francisco, CA, USA; UOFC = University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada; UOFT = University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada; YUHS = Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea.

Adherence to follow-up biopsy schedules

Data on completion and timing of follow-up biopsy were available for 21 of 25 centres. The mean follow-up for all 15 101 men was 28.5 months, with 10 653 having follow-up longer than 1 yr and 2975 longer than 5 yr. Fourteen of these cohorts followed either annual or PRIAS-based schedules, seven in each. This comprised 7224 men with a mean follow-up of 28.3 mo. A total of 5929 (82%) had follow-up after 1 yr and 1448 (20%) after 5 yr (Fig. 2). Adherence to both annual and PRIAS-based re-biopsy schedules declined with increasing time on AS, from 92% and 89% at 1 yr after diagnosis, to 66% and 71% at 7 yr after diagnosis, respectively. Overall odds ratio (OR) for non- adherence with increasing time on AS was 0.23 (95% confidence interval [CI] 0.19–0.28, $p < 0.001$).

Pooled nonadherence rates for annual repeat biopsies increased over time, being 8%, 16%, and 34% at 1, 4, and 7 yr from diagnosis, respectively. Pooled nonadherence rates for PRIAS-based schedules were 11%, 30%, and 29% at 1, 4, and 7 yr after diagnosis, respectively. The generalized linear mixed model revealed no significant difference in nonadherence between cohorts using annual or PRIAS- based biopsy schedules (OR 0.27, 95% CI 0.03–2.54, $p = 0.179$). Fig. 2 shows pooled nonadherence rates by schedule. Biopsies were commonly late in centres performing annual biopsy and early in centres performing 3-yearly biopsy (Supplementary Fig. 1).

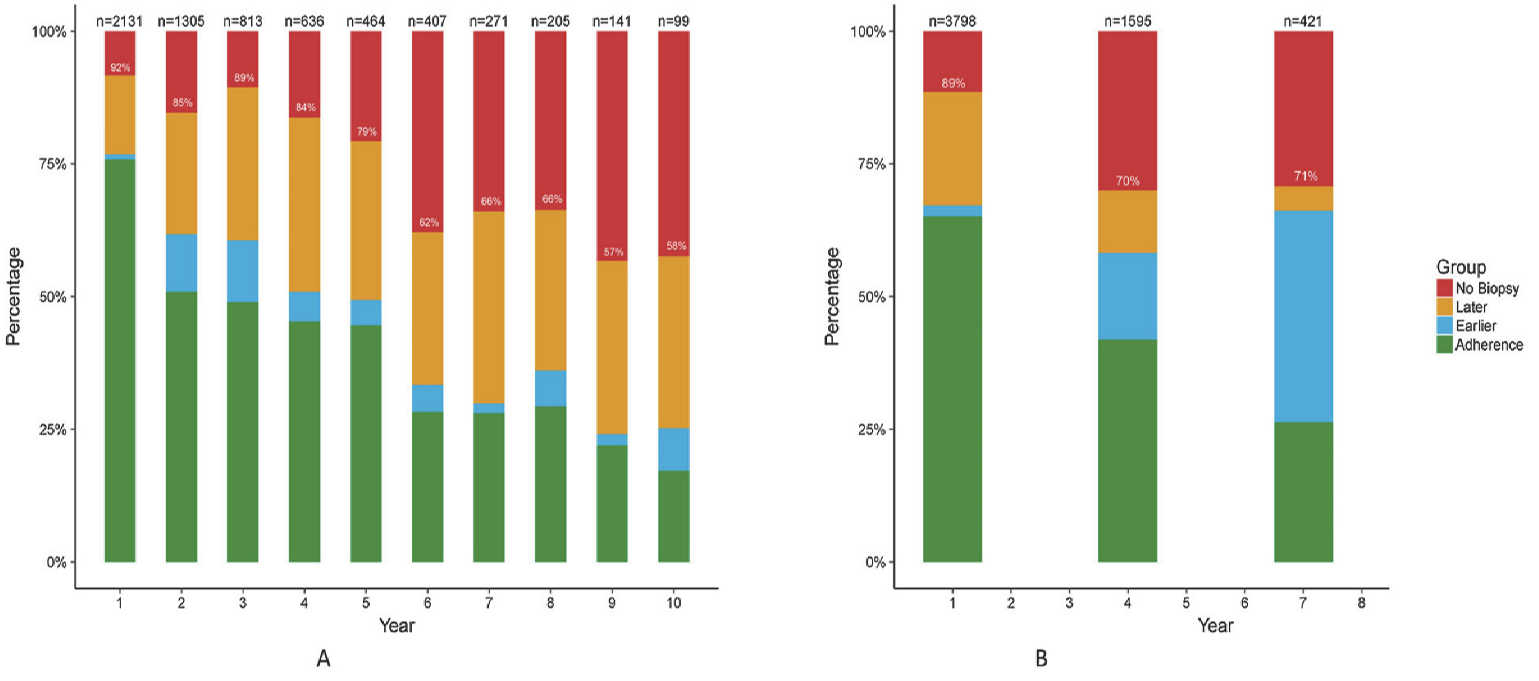


Figure 2 – Pooled rates of adherence to repeat biopsy, by schedule: (A) annual and (B) PRIAS based. Cohorts included in annual group: JHU, MDACC, EU, INT-Milan-SAINT, CHU-Lille, SGH, and KU. Cohorts included in PRIAS-based group: EMC, INT-Milan-PRIAS, UOFT, UBC, HUCH, SUS, and MEASCAP. CHU = Lille University Hospital Center, Lille, France; EMC = Erasmus Medical Center, Rotterdam, the Netherlands; EU = Emory University School of Medicine, Winship Cancer Institute, Atlanta, GA, USA; HUCH = Helsinki University Central Hospital, Helsinki, Finland; INT = Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; JHU = Johns Hopkins University, Baltimore, MD, USA; KU = Kagawa University Faculty of Medicine, Kagawa, Japan; MDACC = MD Anderson Cancer Centre, Houston, TX, USA; MEASCAP = Monash University and Epworth Health, Melbourne, Australia; PRIAS = Prostate Cancer Research International: Active Surveillance; SAINT = Sorveglianza Attiva Istituto Nazionale Tumori; SGH = Singapore General Hospital, Singapore; SUS = Skåne University Hospital, Malmö, Sweden; UBC = University of British Columbia, BC Cancer Agency, Vancouver, Canada; UOFT = University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Canada.

DISCUSSION

AS is an important facet of the management of localized prostate cancer. Although accurate selection of men with low-risk disease and subsequent follow-up is vital to the success of AS, adherence to these parameters has not been investigated thoroughly. In this study, we quantified adherence to site-specific protocols at centres committed to AS by being contributors to the GAP3 consortium.

Adherence to inclusion criteria was either good or excellent across most AS programs; however, slight nonadherence was observed for PSA and GS on diagnosis. This has likely emerged with evidence for expanding criteria in recent years. AS was initially confined to men with PSA $\leq$10 ng/ml and GS 3 + 3. Increasingly, PSA has been identified as an unreliable trigger for intervention [9], and several institutions have broadened their criteria to include men with PSA 10–20 ng/ml and GS 3 + 4 cancers [10]. The four centres with only "good" adherence in our study allowed only GS 3 + 3 and PSA < 10, suggesting that centres without formally expanded criteria may have been reassured in including men outside of regular parameters at diagnosis.

Four of the eight cohorts with fair-weak overall adherence were limited by poor adherence to less commonly used individual criteria of PSAD, number of positive biopsies, and extent of cancer per core. Poor adherence to these volume-based criteria may reflect increasing use of both magnetic resonance imaging (MRI)-guided and transperineal template biopsy techniques. Sampling multiple targeted and systematic cores may inflate the number of positive cores, prompting sites to allow men with higher-volume GS 6 disease despite breaching inclusion criteria.

Evidence for the use of AS in men with intermediate-risk disease remains equivocal. Retrospective studies describe high rates of high-grade and non–organ-confined tumors in men with intermediate-risk disease who underwent upfront radical prostatectomy [11]. Some prospective cohorts have demonstrated results supporting AS in well-selected GS 7 disease [12], although these remain inferior to men with GS 6 tumors [13]. Upfront stratification of these men into favorable and unfavorable intermediate-risk disease is required, along with prospective evaluation of their safety in AS, to justify widespread broadening of the inclusion criteria [14]. Beyond this, further reasons for nonadherence to PSA and GS inclusion criteria need to be explored for forming consensus guidelines to optimize adherence and outcomes.

Data surrounding AS have come largely from the major academic centres with sufficient resources to ensure adequate follow-up, to promptly identify pathological and clinical disease progression. Recent studies, however, have demonstrated that adherence to follow-up protocols in a community setting may be substantially lower [3–5]. Luckenbaugh et al [4] demonstrated only 27% concordance with full AS follow-up protocols in community practices, with the reason for noncompliance being failure to re-biopsy in 82% of cases. Similar results were seen in a recent Australian study, which reported only 26% concordance with follow-up

biopsy and PSA protocols, again mostly due to delay or failure to re-biopsy. This was more prominent in the public, rather than in the private, sector, perhaps reflecting availability of resources [5]. Finally, a large population- based study of more than 5000 men in the USA identified even lower rates of adherence in the community. Less than 13% of men had biopsies after the first 2 yr, and only 11% received follow-up as per PRIAS requirements within the first 5 yr of AS [15].

Even in our series, there was evidence of difficulty in adhering to repeat biopsy protocols. Across all centres, adherence was high in early follow-up but declined with time. Despite heterogeneity with respect to biopsy timing, we found no difference in adherence rates between annual and longer biopsy intervals. However, repeat biopsy was commonly late with high biopsy frequency, and early in protocols with longer intervals between biopsies. Therefore, existing schedules may be altered to define an optimal middle ground. Simulation studies suggest the reduction of repeat biopsies is feasible, with only minor delays in detection of disease progression. Markov model simulations have demonstrated that avoiding six annual biopsies in the first 10 yr of AS increases the risk of detecting grade progression after >24 mo delay by only 10% [16] and that widening of biopsy intervals from yearly to 3 yearly is associated with minimal increase in the risk of cancer- specific mortality [17].

Diminishing adherence to repeat biopsies may be patient driven or clinician driven. Nonadherence may be expected from the perspective of patient experience, although the effect of biopsy morbidity and psychological distress on persistence with AS is unclear [18,19]. Clinically, recent developments such as multiparametric MRI (mpMRI) and genomic markers may be influencing clinical recommendations about continuing surveillance confidently in the absence of repeat prostate biopsies [20]. Hence, a more dynamic and personalized approach to follow-up on AS may address these driving factors for diminishing adherence with increasing time on surveillance [21]. Regardless of the approach, however, it is currently unclear whether frequency of follow-up has an impact on oncological or survival outcomes. Conservative management of localized disease has been associated with high rates of cancer-specific survival at 10 yr [22]; however, this is yet to be evaluated in contemporary AS cohorts.

Our study is not without limitations. Importantly, mpMRI was not routinely used when GAP3 commenced but has since become an increasingly utilised diagnostic tool. Excellent negative predictive values achieved using mpMRI [23] confer a possible reduction in biopsies based on imaging findings alone or in combination with noninvasive clinical parameters [24]. Despite negative predictive values at some centres being as low as 81% [25], mpMRI performs better in determining progression of existing lesions, with 97% of upgraded lesions corresponding with the site of mpMRI abnormality [26]. This may provide part of the solution to the issue of adherence, whereby a Prostate Imaging Reporting and Data System score of 1–2 or a lack of progression of an existing lesion may give confidence to avoid repeat biopsy, accepting a

small risk of missing clinically significant cancer progression. There is early evidence for the success of mpMRI in this setting, particularly when considered in conjunction with PSA density [27].

However, whether surveillance using mpMRI alone is a justifiable solution to the problem of adherence remains uncertain. Targeted mpMRI-guided biopsy detects more clinically significant disease than transrectal ultrasound biopsy [28], but misses up to 20% of significant cancers found on systematic transperineal template biopsy [29]. Likewise, evidence in surveillance populations suggests that significant lesions are commonly present outside the mpMRI-targeted zone, suggesting that systematic biopsy is still required [30]. Beyond mpMRI, emerging tools such as prostate-specific membrane antigen positron emission tomography imaging and genomic biomarkers may provide further individualized reassurance of the safety of continuing surveillance in selected patients, whilst avoiding repeat biopsies.

CONCLUSIONS

We have demonstrated high levels of adherence to site-specific inclusion criteria for AS across centres involved in GAP3 worldwide. Nonadherence to inclusion criteria was primarily associated with PSA and GS, number of positive biopsy cores, and maximum extent of biopsy core involvement. Reasons for this should be further explored and considered in the synthesis of a workable, evidence-based AS selection guideline.

Adherence to biopsy schedules on AS declines over time and is not influenced by biopsy frequency, emphasizing the importance of accurate risk stratification at inclusion. The clinical impact of poor biopsy adherence must be investigated, balancing risk of under-treatment against discomfort associated with repeat biopsy, in constructing a more dynamic and personalized risk-based, rather than fixed, approach to biopsy scheduling on AS.

REFERENCES

1. Tosoian JJ, Mamawala M, Epstein JI, et al. Intermediate and longer- term outcomes from a prospective active-surveillance program for favorable-risk prostate cancer. J Clin Oncol 2015;33:3379–85.
2. Bul M, Zhu X, Valdagni R, etal. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. Eur Urol 2013;63:597–603.
3. Bokhorst LP, Alberts AR, Rannikko A, et al. Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. Eur Urol 2015;68:814–21.
4. Luckenbaugh AN, Auffenberg GB, Hawken SR, et al. Variation in guideline concordant active surveillance followup in diverse urology practices. J Urol 2017;197(3 Pt 1):621–6.
5. Evans MA, Millar JL, Earnest A, et al. Active surveillance of men with low risk prostate cancer: evidence from the Prostate Cancer Outcomes Registry. Victoria. Med J Aust 2018;208:439–43.
6. Hulsen T, Obbink H, Van Der Linden W, et al. 958 Integrating large datasets for the Movember Global Action Plan on active surveillance for low risk prostate cancer. Eur Urol Suppl 2016;15:e958.
7. Bruinsma SM, Zhang L, Roobol MJ, et al. The Movember Foundation's GAP3 cohort: a profile of the largest global prostate cancer active surveillance database to date. BJU Int 2018;121:737–44.
8. Van Hemelrijck M, Ji X, Helleman J, et al. Reasons for discontinuing active surveillance assessment of 21 centres in 12 countries in the Movember GAP3 Consortium. Eur Urol 2019;75:523–31.
9. Ross AE, Loeb S, Landis P, et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. J Clin Oncol 2010;28:2810–6.
10. Loeb S, Bruinsma SM, Nicholson J, et al. Active surveillance for prostate cancer: a systematic review of clinicopathologic variables and biomarkers for risk stratification. Eur Urol2015;67:619-26.
11. Morlacco A, Cheville JC, et al . Adverse disease features in Gleason score 3 + 4 "favorable intermediate-risk" prostate cancer: implications for active surveillance. Eur Urol 2017;72:442–7.
12. Nyame YA, Almassi N, Haywood SC, et al. Intermediate term outcomes for men with very low/low/intermediate/high risk prostate cancer managed by active surveillance. J Urol 2017;198:591–9.
13. Musunuru HB, Yamamoto T, Klotz L, et al. Active surveillance for intermediate risk prostate cancer: survival outcomes in the Sunny- brook experience. J Urol 2016;196:1651–8.
14. Verbeek JFM, Roobol MJ, Steyerberg EW. Risk-based selection for active surveillance. In: Klotz L, editor. Active surveillance for localized prostate cancer: a new paradigm for clinical management. ed. 2. Cham, Switzerland: Springer International Publishing; 2018. p. 53–64.
15. Loeb S, Walter D, Curnyn C, et al. How active is active surveillance? Intensity of follow up during active surveillance for prostate cancer in the United States. J Urol 2016;196:721–6.
16. Barnett CL, Auffenberg GB, Cheng Z, et al. Optimizing active surveillance strategies to balance the competing goals of early detection of grade progression and minimizing harm from biopsies. Cancer 2018;124:698–705.
17. de Carvalho TM, Heijnsdijk EAM, de Koning HJ. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. BJU Int 2017;119:560–6.
18. Egger SJ, Calopedos RJ, O'Connell DL, el al. Long-term psychological and quality-of-life effects of active surveillance and watchful waiting after diagnosis of low-risk localized prostate cancer. Eur Urol 2018;73:859–67.

19. Tan H-J, Marks LS, Hoyt MA, et al. The relationship between intolerance of uncertainty and anxiety in men on active surveillance for prostate cancer. J Urol 2016;195:1724–30.

20. Shore N, Eure G, Saltzstein D, et al. MP35-14 A 17-gene assay drives high active surveillance persistence in clinically low-risk prostate cancer: 1-year results from a 1,200-patient prospective observational trial. J Urol 2018;199(April):e452.

21. Nieboer D, Tomer A, Rizopoulos D, Roobol MJ, Steyerberg EW. Active surveillance: a review of risk-based, dynamic monitoring. Transl Androl Urol 2018;7:106–15.

22. Adolfsson J, Rönström L, Löwhagen T, Carstensen J, Hedlund PO. Deferred treatment of clinically localized low-grade prostate cancer: the experience from a prospective series at the Karolinska Hospital. J Urol 1994;152(5 Pt 2):1757–60.

23. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389:815–22.

24. Radtke JP, Wiesenfarth M, Kesch C, etal. Combined clinical parameters and multiparametric magnetic resonance imaging for advanced risk modeling of prostate cancer-patient-tailored risk stratification can reduce unnecessary biopsies. Eur Urol 2017;72:888–96.

25. Frye TP, George AK, Kilchevsky A, et al. Magnetic resonance imaging- transrectal ultrasound guided fusion biopsy to detect progression in patients with existing lesions on active surveillance for low and intermediate risk prostate cancer. J Urol 2017;197(3 Pt 1):640–6.

26. Nassiri N, Margolis DJ, Natarajan S, et al. Targeted biopsy to detect Gleason score upgrading during active surveillance for men with low versus intermediate risk prostate cancer. J Urol 2017;197(3 Pt 1):632–9.

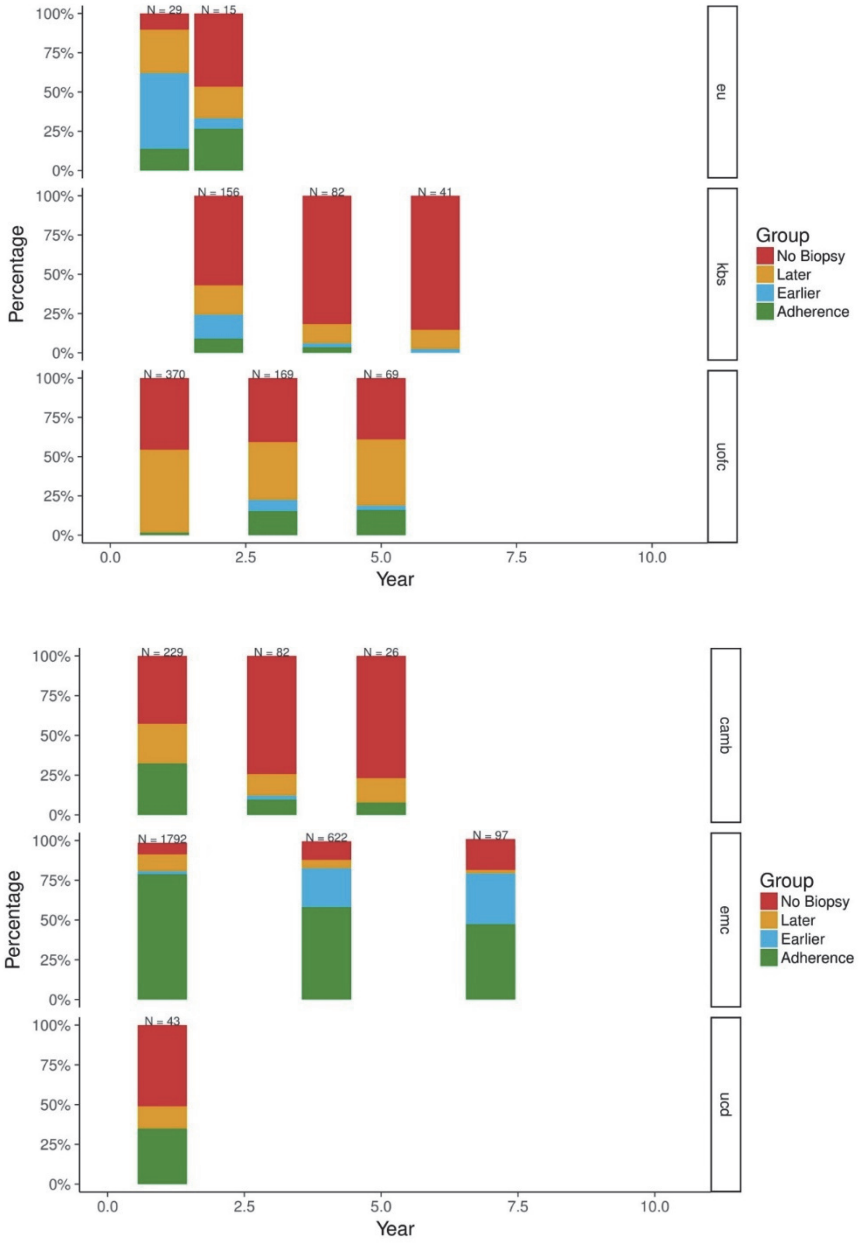
27. Gallagher KM, Christopher E, Cameron AJ, et al. Four-year outcomes from a multiparametric magnetic resonance imaging (MRI)-based active surveillance programme: PSA dynamics and serial MRI scans allow omission of protocol biopsies. BJU Int 2019;123:429–38.

28. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018;378:1767–77.

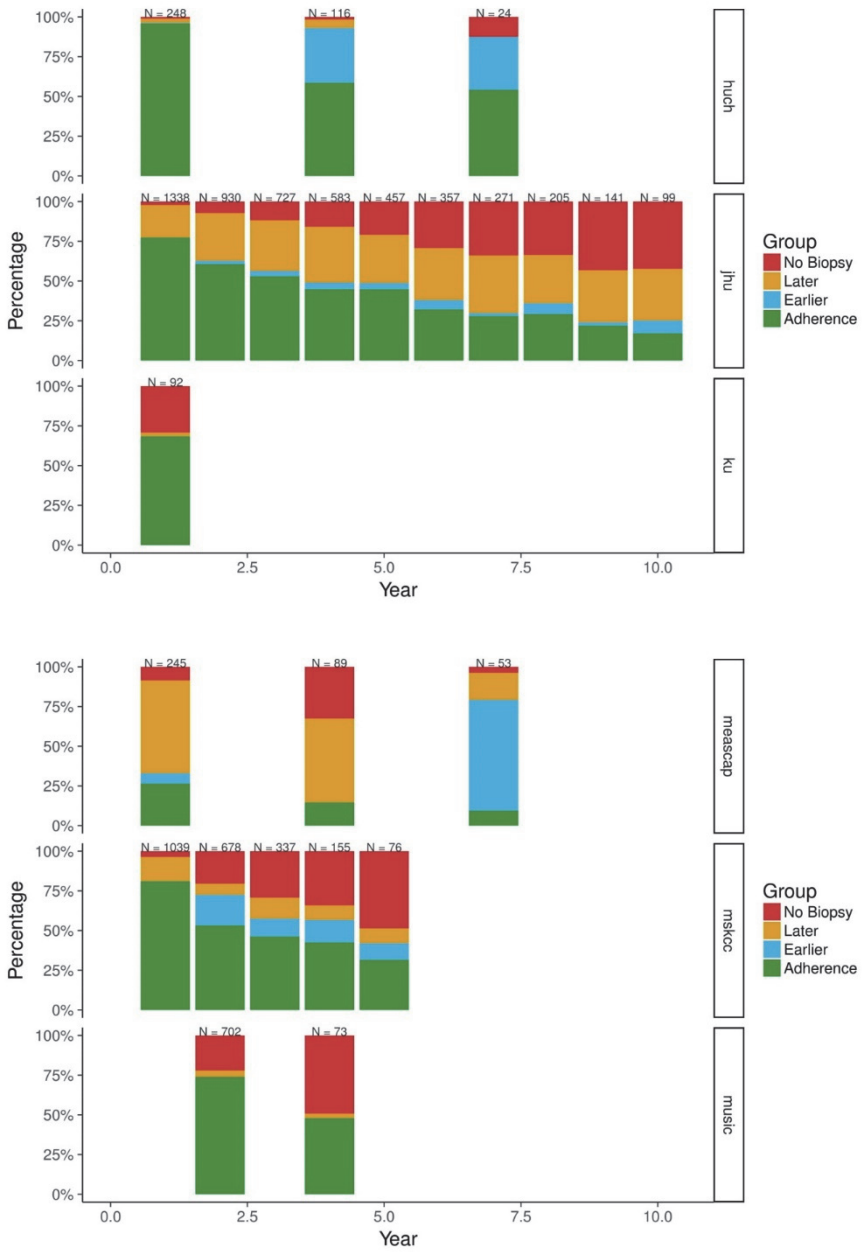
29. Mortezavi A, Märzendorfer O, Donati OF, et al. Diagnostic accuracy of multiparametric magnetic resonance imaging and fusion guided targeted biopsy evaluated by transperineal template saturation prostate biopsy for the detection and characterization of prostate cancer. J Urol 2018;200:309–18.

30. Tran GN, Leapman MS, Nguyen HG, et al. Magnetic resonance imaging-ultrasound fusion biopsy during prostate cancer active surveillance. Eur Urol 2017;72:275–81.

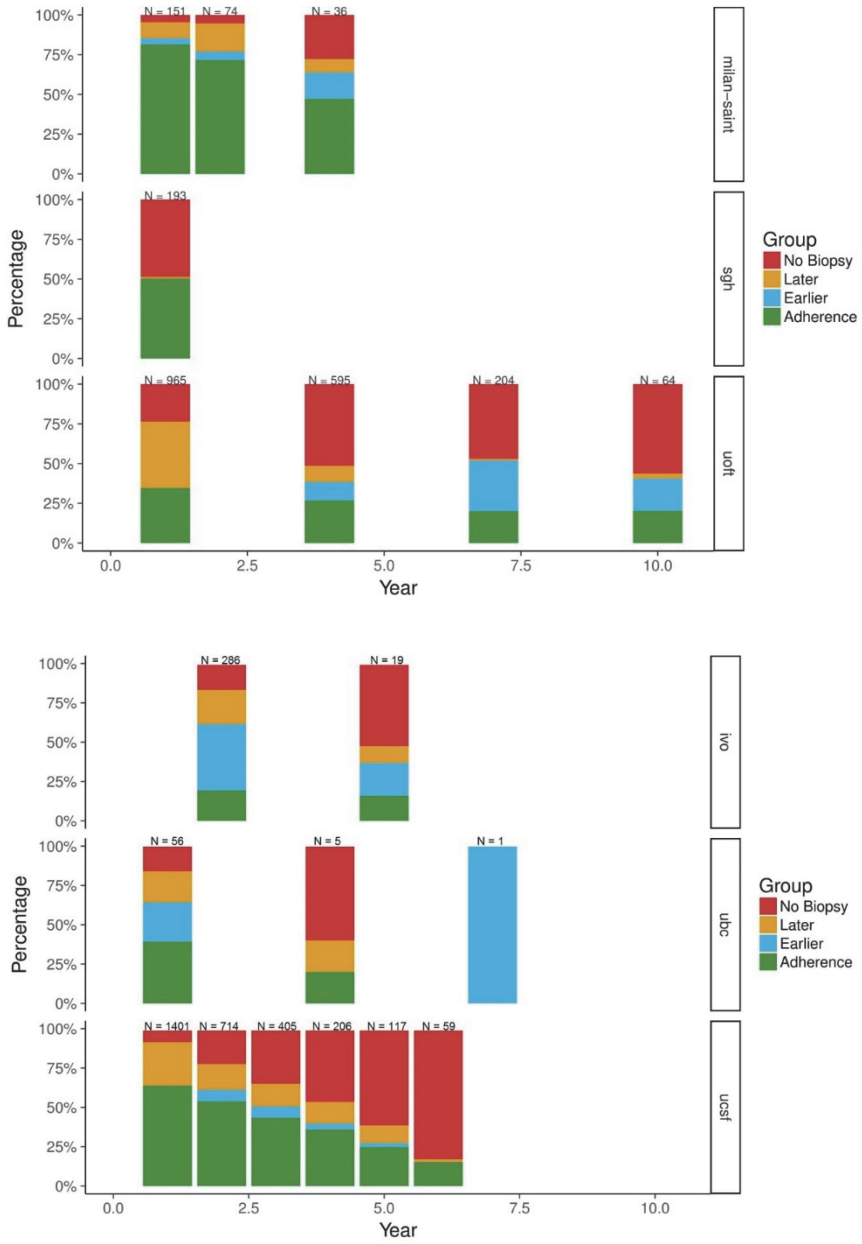
SUPPLEMENTARY MATERIALS



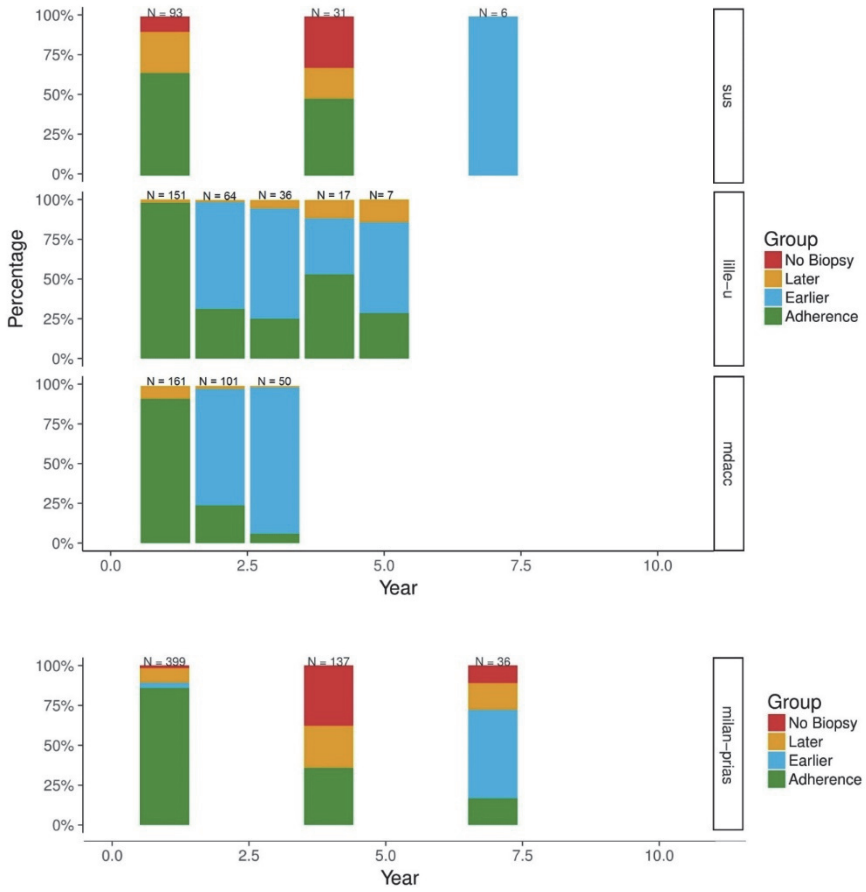
Supplementary Figure 1: Adherence plots for re-biopsy schedules, by centre.



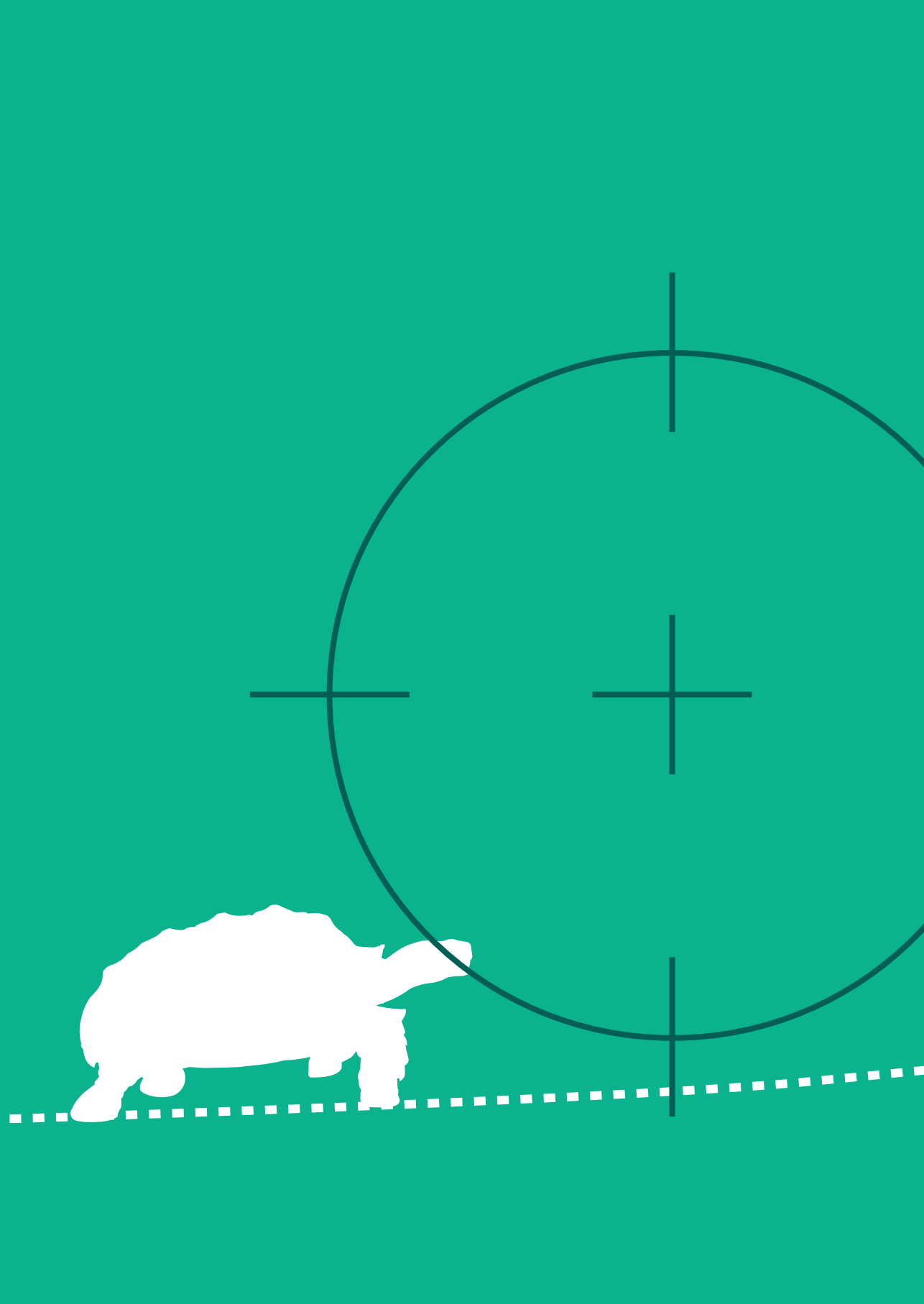
Supplementary Figure 1: Adherence plots for re-biopsy schedules, by centre.



Supplementary Figure 1: Adherence plots for re-biopsy schedules, by centre. Note that the x-axis for UCSF is biopsy number, rather than years, as biopsy frequency varies per patient according to this centre's active surveillance protocol.



Supplementary Figure 1: Adherence plots for re-biopsy schedules, by centre.



CHAPTER 4

Reporting and interpreting decision curve analysis: a guide for investigators

Van Calster B*
Wynants L*
Verbeek JFM
Verbakel JY
Christodoulue E
Vickers AJ
Roobol MJ
Steyerberg EW

* These authors are joint first authors

Eur Urol. 2018 Dec;74(6):796-804

ABSTRACT

Context

Urologists regularly develop clinical risk prediction models to support clinical decisions. In contrast to traditional performance measures, decision curve analysis (DCA) can assess the utility of models for decision making. DCA plots net benefit (NB) at a range of clinically reasonable risk thresholds.

Objective

To provide recommendations on interpreting and reporting DCA when evaluating prediction models.

Evidence acquisition

We informally reviewed the urological literature to determine investigators' understanding of DCA. To illustrate, we use data from 3616 patients to develop risk models for high-grade prostate cancer (n=313, 9%) to decide who should undergo a biopsy. The baseline model includes prostate-specific antigen and digital rectal examination; the extended model adds two predictors based on transrectal ultrasound (TRUS).

Evidence synthesis

We explain risk thresholds, NB, default strategies (treat all, treat no one), and test tradeoff. To use DCA, first determine whether a model is superior to all other strategies across the range of reasonable risk thresholds. If so, that model appears to improve decisions irrespective of threshold. Second, consider if there are important extra costs to using the model. If so, obtain the test tradeoff to check whether the increase in NB versus the best other strategy is worth the additional cost. In our case study, addition of TRUS improved NB by 0.0114, equivalent to 1.1 more detected high-grade prostate cancers per 100 patients. Hence, adding TRUS would be worthwhile if we accept subjecting 88 patients to TRUS to find one additional high-grade prostate cancer or, alternatively, subjecting 10 patients to TRUS to avoid one unnecessary biopsy.

Conclusions

The proposed guidelines can help researchers understand DCA and improve application and reporting.

Patient summary

Decision curve analysis can identify risk models that can help us make better clinical decisions. We illustrate appropriate reporting and interpretation of decision curve analysis.

INTRODUCTION

Clinical risk prediction models are commonly developed in urology and other medical fields to predict the probability or risk of a current disease (eg, biopsy-detectable aggressive prostate cancer), or a future state (eg, cancer recurrence) [1-3]. Such models are usually evaluated with statistical measures for discrimination and calibration. Discrimination evaluates how well the predicted risks distinguish between patients with and without disease. The c-statistic is the most commonly used measure for discrimination. Calibration evaluates the reliability of the estimated risks: if we predict 10%, on average 10 out of 100 patients should have the disease [1], [4]. Assessments of calibration may include graphs and statistics such as observed versus expected ratios or calibration slopes. Although a model with better discrimination and calibration should theoretically be a better guide to clinical management [4-6], statistical measures fall short when we want to evaluate whether the risk model improves clinical decision making. Such measures cannot inform us whether it is beneficial to use a model to make clinical decisions or which of two models leads to better decisions, especially if one model has better discrimination and the other better calibration [7].

To overcome this limitation, decision-analytic measures have been developed to summarize the performance of the model in supporting decision making. We focus on net benefit (NB) as the key part of decision curve analysis (DCA), which was introduced in 2006 [8]. Editorials supporting DCA have been published in leading medical journals including *JAMA, Lancet Oncology, Journal of Clinical Oncology, BMJ, PLoS Medicine, and Annals of Internal Medicine* [9-17]. Importantly, evaluating NB is recommended by the TRIPOD guidelines for prediction models [18]. DCA is widely used within urology and many other clinical fields. A Web of Science search (September 11, 2018) revealed that the 2006 paper was cited 703 times in total. DCA was most often cited in journals from urology and nephrology (176 citations), oncology (147), and general and internal medicine (76). European Urology is the journal with most citations (45).

However, based on various personal discussions, we notice that researchers struggle with the interpretation and reporting of NB. We therefore aim to provide an investigators' guide to NB and DCA. A case study on prediction of high-grade prostate cancer is used as an illustrative example.

EVIDENCE ACQUISITION

We informally reviewed the urological literature to determine investigators' understanding of DCA. To illustrate, we use data from 3616 patients to develop risk models for high-grade prostate cancer (n = 313, 9%) to decide who should undergo a biopsy. The baseline model includes prostate-specific antigen (PSA) and digital rectal examination; the extended model adds two predictors based on transrectal ultrasound (TRUS).

EVIDENCE SYNTHESIS

Case study: prediction of high-grade prostate cancer to decide who to biopsy

Screening with PSA results in overdiagnosis of indolent prostate cancer [19]. Risk calculators have been developed for high-grade prostate cancer [20]. Using these models to decide who to biopsy can reduce unnecessary biopsies, which are aversive procedures with a risk of sepsis and lead to detection of indolent disease. Detecting high-grade prostate cancer is important, because early detection of these potentially lethal cancers can lead to curative treatment [21].

The Rotterdam Prostate Cancer Risk Calculator (RPCRC) predicts the risk of high-grade cancer in an individual patient based on PSA, abnormal digital rectal examination (DRE), abnormal TRUS findings, and TRUS-based prostate volume [22]. The RPCRC was developed from the European Randomized Study of Screening for Prostate Cancer, Rotterdam section. Men between ages 54 and 74 yr and with PSA ≥3.0 ng/ml received lateral sextant biopsy between November 1993 and March 2000 (n = 3616). The outcome was high-grade prostate cancer (n = 313, 9%), defined as Gleason score 3 + 4 or higher on biopsy and/or tumor stage >T2b.

We focused on a baseline model containing two predictors: PSA value and abnormal DRE. Then we fitted an extended model to investigate the additional value of abnormal TRUS and TRUS-based prostate volume (Table 1). We had 313 events for two or four model coefficients (ie, at least 78 events per variable), substantially limiting the risk of overfitting. To check this, we calculated the calibration slope using bootstrapping [1], [4]. Where a slope of 1 indicates no overfitting, we found slopes of 0.998 for the baseline model and 0.995 for the extended model, suggesting a marginal 0.2–0.5% overfitting. The c-statistic was 0.814 (95% confidence interval 0.785–0.840) for the baseline model and 0.866 (0.841–0.888) for the extended model. Thus, based on the traditional metrics of discrimination and calibration, most researchers would agree that the extended model is clearly better.

Table 1. Baseline and extended models to predict high-grade prostate cancer.

Predictor	Median (IQR) or *n* (%)	Baseline model		Extended model	
		***B* (SE)**	**OR (95% CI)**	***B* (SE)**	**OR (95% CI)**
Intercept		−5.68 (0.21)		−0.20 (0.67)	
PSA[a]	4.3 ng/ml (3.1–6.4)	1.03 (0.063)	2.79 per doubling (2.47–3.16)	1.21 (0.072)	3.36 per doubling (2.92–3.87)
Abnormal DRE	1279 (35%)	1.60 (0.14)	4.95 (3.79–6.46)	1.03 (0.15)	2.81 (2.10–3.76)
Abnormal TRUS	1229 (34%)			1.21 (0.15)	3.35 (2.50–4.48)
Tumor volume[a]	41 ml (32–55)			−1.16 (0.13)	0.31 per doubling (0.24–0.41)

B = regression coefficient; CI = confidence interval; DRE = digital rectal examination; IQR = interquartile range; OR = odds ratio; PSA = prostate-specific antigen; SE = standard error; TRUS = transrectal ultrasound. [a] PSA and tumor volume are modeled with log2 transformation, such that the odds ratios for these variables represent the change in odds per doubling of the PSA/volume.

Risk thresholds

To use a risk model for treatment decisions, we specify a risk threshold T above which we would treat. In our example, treatment refers to biopsy; however, depending on the application, "treatment" can refer to a wide range of interventions, such as additional diagnostic workup, referral to specialized care, a procedure (eg, lymph node resection), delaying surgery (eg, in patients at high risk of complications), medical treatment, or lifestyle changes. In our prostate biopsy example, we could recommend biopsy if the predicted risk of high-grade cancer was 10% or more (T = 10%) and otherwise advise monitoring without biopsy. Correct classifications are labeled true positives (for patients with the event) or true negatives (for patients without the event). Incorrect decisions are labeled false negatives and false positives.

Many investigators select a threshold that maximizes the sum of the true positive and true negative rates [23]. However, this assumes that sensitivity and specificity are equally important. Relevant thresholds incorporate clinical considerations for decision making. In our case, it is more important to find an aggressive cancer than to avoid unnecessary biopsy. According to decision theory, the risk threshold reflects the risk at which we are indifferent about treatment [24]. Assume that we are willing to biopsy no more than 10 men in order to find one high-grade prostate cancer. Then we consider the benefit of detecting one high-grade prostate cancer to be nine times larger than the harm of an unnecessary biopsy: the "harm-to-benefit" ratio is 1:9. This ratio is hard to specify directly. Fortunately, it has a direct

relationship with the risk threshold T: the odds of T equal the harm-to-benefit ratio [24]. For example, a risk threshold of 10% implies a harm-to-benefit ratio of 1:9 (odds [10%] = 10/90).

A reasonable risk threshold for decision-making involves a holistic assessment of all possible outcomes. A biopsy can be painful and inconvenient, and entails a risk of infection; therefore, it is preferable to avoid a biopsy when deemed unnecessary. In case the patient has high-grade prostate cancer, the biopsy can lead to cancer treatment, which may improve prognosis but may cause side effects. Hence, different strategies have their benefits and harms, which may also be of financial or organizational nature. Balancing of all benefits and harms determines which risk thresholds are reasonable.

NET BENEFIT & DECISION CURVE ANALYSIS

The utility of risk models may be evaluated with cost-effectiveness studies [25], supported by empirical evaluations of the impact of using a model in clinical practice. Such studies are difficult to conduct. Instead, there are simpler measures to evaluate the potential clinical utility of risk models [26]. We focus on NB, which combines the number of true positives and false positives into a single "net" number [8], [9]. NB is a concept similar to that of net profit in business: income minus expenditure. In the prostate cancer example, the "income" represents true positives—cases of aggressive prostate cancer found early; the "expenditure" represents false positives—unnecessary biopsies. In most medical scenarios, income and expenditure are on different scales. Therefore, we need an "exchange rate" to reflect the balance between the benefit of a true positive and the harm of a false positive (the harm-to-benefit ratio). Going back to our example, with a risk threshold of 10%, we would weigh each false positive by the odds of 10 ($10 \div 90 = 0.1111$). The baseline model at T = 10% yields 211 detected high-grade prostate cancers and 621 unnecessary biopsies. Then, 211 true positives minus $(10 \div 90) \times 621$ false positives give 142 "net" true positives. Correction for the harm of the unnecessary biopsies adjusts the observed 211 detected high-grade prostate cancers to a net number of 142. The net result is positive because there were only 2.9 false positives per true positive ($= 621 \div 211$) at the 10% risk threshold, whereas this threshold implies that we are willing to accept much more unnecessary biopsies (ie, nine) per detected high-grade prostate cancer. NB is obtained by dividing the net true positives by the sample size, which gives 0.0393 for the baseline model (Table 2). This means that there are 3.9 net detected high-grade prostate cancers per 100 patients. The division by sample size avoids that the magnitude of NB depends on the size of a dataset. Several measures have been proposed that are closely related to NB and that lead to identical conclusions (see the Supplementary materials) [16], [26-28]. Usually, there is no single risk threshold that is universally acceptable and so it is important to evaluate NB over a range of reasonable thresholds [9], [29]. In the case of prostate biopsy, for example, a patient averse to the risk of

untreated cancer may prefer a lower risk threshold, whereas a patient less tolerant of invasive procedures such as biopsy may choose a higher threshold. The clinical decision for which the model is used is pivotal to set the relevant threshold range. For example, using a risk model to select patients with suspicious bladder tumors for general urological surgery will require a different threshold compared with using the model to select patients for specialized oncological surgery. A decision curve plots NB for a range of relevant risk thresholds (Fig. 1). In our example, we focused on thresholds between 5% and 20%.

Table 2. Net benefit and test tradeoff results for the baseline and extended models to predict high-grade prostate cancer at a risk threshold of 10%.

Statistic	Result
Default strategies	
NB if all men subject to biopsy (NB_{TrA})	−0.0149
NB if no one subject to biopsy (NB_{TrN})	0
Baseline model	
NB if baseline model is used to select patients for biopsy	0.0393
Detected HG-PCa without unnecessary biopsies	3.9 per 100 patients
Test tradeoff, patients biopsied per detected HG-PCa	25.4
Extended versus baseline model	
NB if extended model is used to select patients for biopsy	0.0507
NB difference between extended and baseline models	0.0114
Additional HG-PCa detected (without change in unnecessary biopsies) when using the extended model rather than the baseline model	1.14 per 100 patients
Test tradeoff, patients undergoing TRUS per additionally detected HG-PCa	87.7
Test tradeoff, patients undergoing TRUS per avoided unnecessary biopsy	9.7

HG-PCa = high-grade prostate cancer; NB = net benefit; TrA = treat all; TrN = treat none; TRUS = transrectal ultrasound.

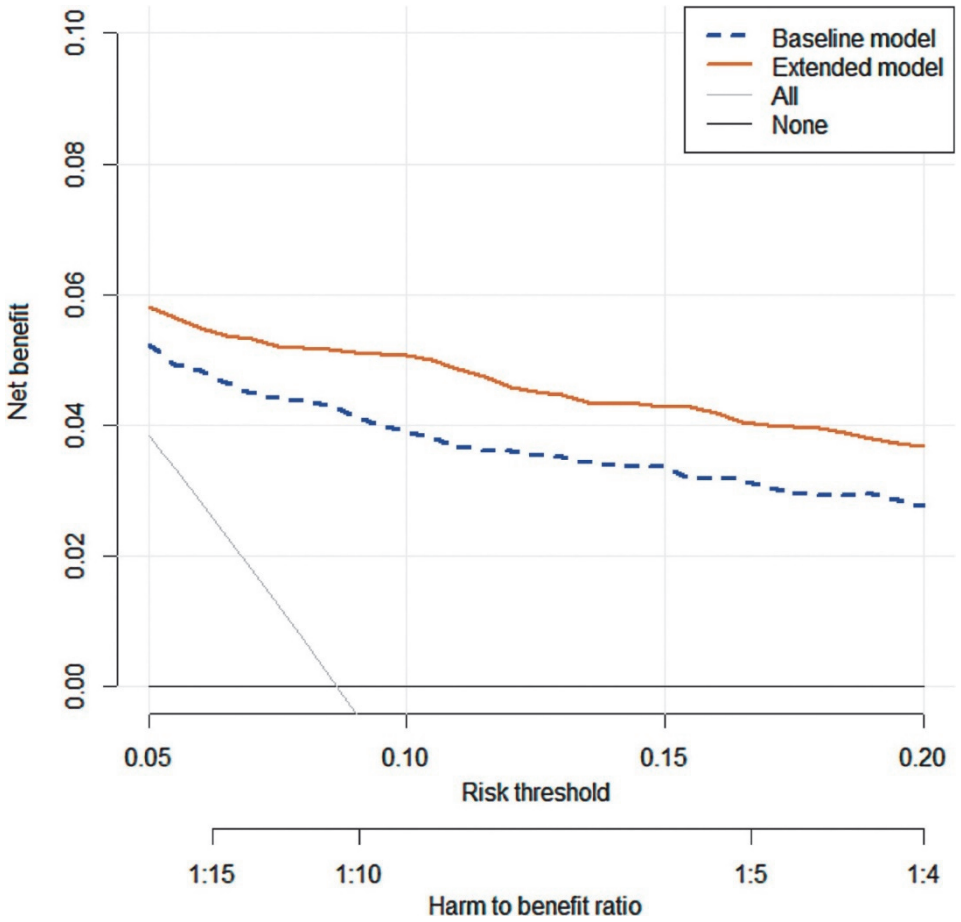


Figure 1. Decision curves for the default strategies and for the baseline and extended models.

Are model-based decisions useful? Comparison with default strategies

To interpret NB properly, we introduce two default strategies where patients are managed without the use of a model [8]. We can biopsy either all patients ("treat all") or no one ("treat none"). NB of treat none is always 0 because this strategy has no true or false positives. Treat all does not imply that T has been set to 0. Rather, the decision to treat everyone is evaluated at all reasonable values of T (see the Supplementary materials for formula). For risk thresholds below prevalence, treat all has a higher NB than treat none. For thresholds above prevalence, the opposite holds true, which implies a negative NB for treat all. At the 10% risk threshold, treat all has an NB of −0.0149.

A model is only clinically useful at threshold T if it has a higher NB than treat all and treat none. If a model has a lower NB than any default strategy, we consider the model clinically harmful:

one of the default strategies leads to better decisions. Importantly, when models are calibrated, they cannot be harmful [4], [5]. Only miscalibrated models can be harmful. For example, if we underestimate the risk of high-grade prostate cancer, we would too often advise against biopsy, missing more cancers than anticipated, leading to poorer NB. When applying DCA, we first evaluate whether the model(s) under study has (have) a higher NB than the default strategies. When comparing two models, we check which model has the highest NB. When one of the models is harmful, further model comparison is redundant. The baseline and extended models of our case study outperform the default strategies across the relevant threshold range, and the extended model outperforms the baseline model.

To interpret DCA results, we illustrate various hypothetical scenarios in Figure 2. We show decision curves for an application where the threshold probability is typically about 20%, but a reasonable range of thresholds is determined to be 10–30%. We show threshold probabilities outside this range for didactic purposes. In Figure 2A, the model (dashed line) has a higher NB than both treat all (full line) and treat none (dotted line) only for threshold probabilities above 20%. As the range of reasonable thresholds is 10–30%, that is, some patients would choose treatment if their risk was only 10% or 15%, the model is not of value. Indeed, for patients with these types of thresholds, NB of the model is worse than the strategy of "treating all", that is, opting for treatment irrespective of the risk from the model. The lower NB at these thresholds is because the model is miscalibrated, slightly underestimating the risk. In Figure 2B, we show a well-calibrated model with a relatively high area under the curve. However, the prevalence of disease in the study is very high (~60%). With baseline risk being very high, it is very difficult for a model to push the risk low enough for a patient to refuse treatment. The model has a higher NB for only a small part of the range of reasonable thresholds, and therefore the model is not of value. In Figure 2C, the model is of benefit for almost, but not quite, the whole of the reasonable range 10–30%: the curves diverge only at the threshold probability of about 13%. However, NB of the model is about the same as the NB of treat all below 13%. Therefore, if the investigators believed that it was not common to have such low threshold probabilities, they could probably justify clinical use. In Figure 2D, either the model or the competing binary test (dashdotted line) has a higher NB than treating all or no patients across the entire range of reasonable threshold probabilities. However, the curves cross in the middle of the reasonable range. In general, the conclusion would be that no strategy is optimal across the whole range of reasonable threshold probabilities, and hence further research is required. However, depending on the clinical situation, there might be calls to favor the model or the test. For instance, NB for each is similar in the key range of thresholds, so if one approach is superior in terms of costs and risks of convenience, then that might be the approach chosen. In Figure 2E, the model (dashed line) is well calibrated with a c-statistic of 0.75. The competing model (dashdotted line) has a c-statistic of 0.80 but is miscalibrated (risks are underestimated). As a result, the model with the lowest c-statistic

is superior in the entire reasonable range of threshold probabilities. The miscalibrated model is even harmful for thresholds up to 15%.

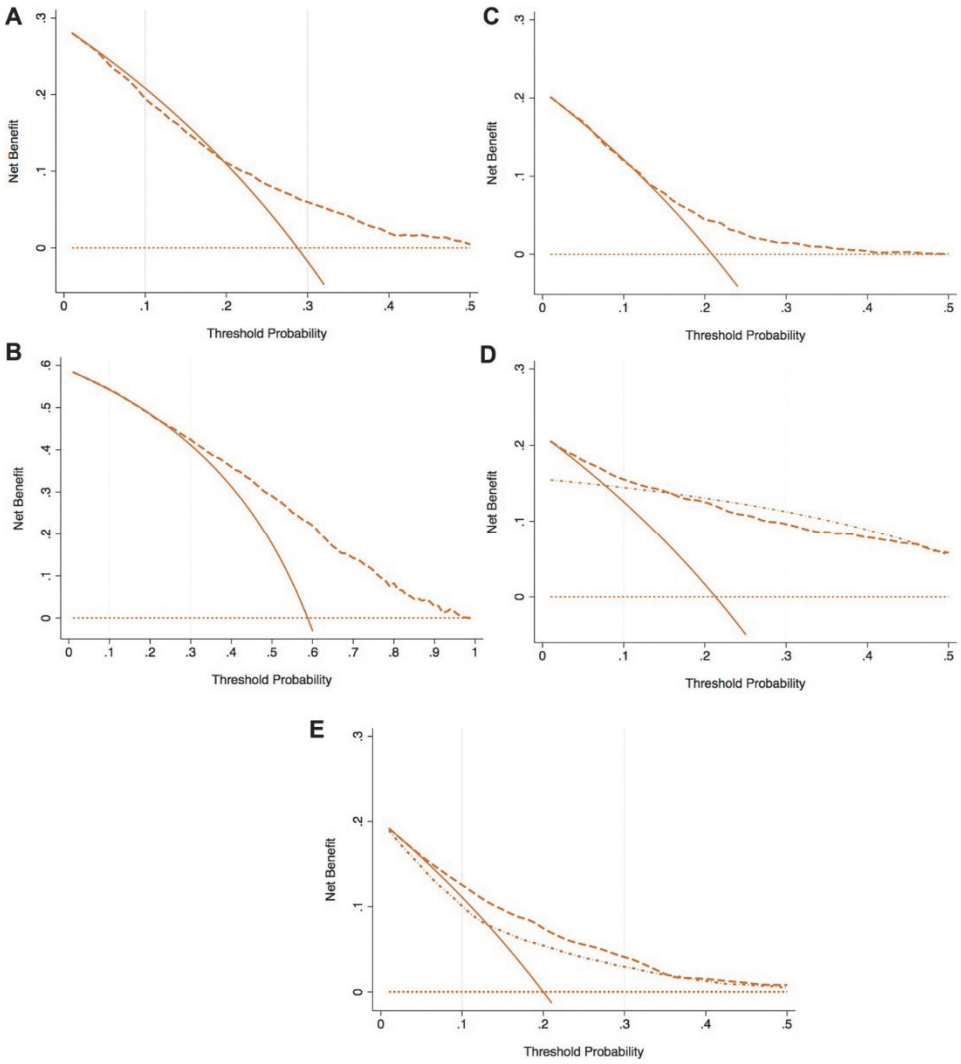


Figure 2. Hypothetical decision curves illustrating several possible scenarios. Panels A, B, C: decision curve analysis for a single model; Panels D and E: decision curve analysis for two competing models. Full line: treat all, dotted line: treat none, dashed line: model, dash dotted line: competing model/test.

Interpretation of NB

NB gives the proportion of "net" true positives in the dataset: the observed number of true positives is corrected for the observed proportion of false positives weighted by the odds of the risk threshold, and the result is divided by the sample size. This "net" proportion is equivalent to the proportion of true positives in the absence of false positives (ie, perfect specificity). The baseline model has an NB of 0.0393 at the 10% risk threshold, which is equivalent to detecting 3.93 (≈4) high-grade prostate cancers and suggesting zero unnecessary biopsies per 100 patients (ie, four true positives and zero false positives). In fact, this is a direct comparison with treat none, which has zero true positives and zero false positives by default. Even though a model may compare well with treat none (ie, NB is positive), it may still be worse than treat all. This is possible when the risk threshold is below prevalence, because then the NB of treat all is higher than the NB of treat none.

To interpret the NB difference between models, consider that the extended model yielded 236 true positives and 475 false positives at the 10% risk threshold (NB = 0.0507). The difference in NB for the extended versus baseline model is 0.0507–0.0393 = 0.0114. The extended model has 1.14 more net detected high-grade prostate cancers per 100 patients. This is equivalent to having 1.14 more detected high-grade prostate cancers per 100 patients for the same number of unnecessary biopsies.

Test tradeoff

NB does not directly account for the cost and harms associated with measuring the predictors in the model. This is a reasonable assumption where the model includes only routinely available data (such as in our base model of PSA and DRE), but if a predictor requires an invasive or expensive test (such as TRUS), we should account for the harm or cost of measurement. We may specify the harms of a model upfront: we ask clinicians "how many of these tests would you do to find one case (eg, high-grade prostate cancer) if the test were 100% perfect"; the reciprocal of that number is the "test harm," which is subtracted from NB [8]. Test harm may be difficult to specify directly. Alternatively, we can focus on the difference in NB (ΔNB) to derive the "test tradeoff" [30-32].

Evaluation of a single model - When validating a single model, ΔNB refers to the difference between the NB of the model and the NB of the best default strategy. The test tradeoff, 1/ΔNB, is the minimum number of tests per true positive that we have to accept to make the model worthwhile given its cost. For the baseline model at 10%, ΔNB is 0.0393 and the test tradeoff is 25.4 (=1/0.0393). If we are willing to use the baseline model on 25 patients to detect one high-grade prostate cancer, this model is worthwhile.

Model comparison - The test tradeoff for the comparison of two models refers to the minimum number of tests for one additional true positive with the best model to make this model

worthwhile given its additional cost. At the 10% risk threshold, ΔNB of the extended versus the baseline model is 0.0114, and the test tradeoff is 87.7 (=1/0.114). If we consider it acceptable to subject 88 patients to TRUS to detect one additional high-grade prostate cancer compared with the model without TRUS, the utility of the extended model is worth the cost of TRUS.

Test tradeoff in terms of true negatives - NB is based on the numbers of true and false positives. From these numbers, it is easy to derive the numbers of true and false negatives. It is therefore possible to obtain an alternative expression of NB, which corrects the number of true negatives for the weighted number of false negatives (see the formula in the Supplementary materials) [33]. As a result, we can express the test tradeoff in terms of true negatives as well. This test tradeoff, obtained as odds (T)/ΔNB, gives the number of patients we should be willing to classify with the best model per additional true negative.

For evaluation of a single model (the baseline model for the case study), the test tradeoff in terms of true negatives equals 2.8 (=odds [10%]/0.0393). If we are willing to use the baseline model on three patients to avoid one unnecessary biopsy, this model is worthwhile. When comparing the extended model with the baseline model at 10%, we find a test tradeoff of 9.7 patients per additional true negative (=odds [10%]/0.0114). The extended model is preferable over the baseline model if we accept doing TRUS on 10 patients to avoid one additional unnecessary biopsy.

Interpretation of the test tradeoffs for the case study - When evaluating the baseline model with the default strategies, the test tradeoff indicates that the baseline model is clearly of value given that the model only requires data that the urologist already has at hand.
TRUS could be invasive and unpleasant; hence, the test tradeoffs can be considered high. Some urologists would not agree to subject 88 patients to TRUS to find one high-grade cancer or perform 10 TRUS to avoid one biopsy (Table 2), despite an increase in the c-statistic of 0.052. Other urologists may accept the test tradeoffs given that TRUS has almost no complications. Nevertheless, we might consider alternative sources, such as magnetic resonance imaging or DRE, to measure volume [34-36].

Recommendations for practice

Interpreting the results of NB - The first step in DCA is to determine whether any model is superior to all other models, and the default strategies of treating all or no patient, across the full range of reasonable threshold probabilities. If so, we can declare that the use of that model would improve patient outcome irrespective of patient or doctor preference. The second step is to consider whether there are important risks, harms, or costs to using the model. If so, we need to interpret the magnitude of the increase in NB versus best default or the competing model, and evaluate the test tradeoff, or use test harm, to check whether the increase in NB is worth the additional cost and harm of using the model.

Defining the treatment decision clearly - DCA evaluates the utility of a model or test to decide who should receive treatment, which can be any diagnostic or therapeutic intervention depending on the application. It is therefore important to unambiguously define the decision. If a model serves mainly to counsel patients (eg, survival probabilities), the meaning of decision curves becomes debatable since the range of personal decisions is wide. For example, a model predicting probability of death at 1 yr in patients with advanced cancer might be used to inform decisions ranging from sorting out legal affairs to travel plans or retirement.

Defining a reasonable range of risk threshold - For a particular treatment decision, utility should be evaluated for a reasonable range of thresholds only. "Reasonable" means that no one would reasonably use a threshold outside that range to decide upon treatment. We therefore recommend showing and interpreting decision curves only for the adopted reasonable range. The ideal situation is when one model shows the highest utility over the entire range. Elsewhere, we have given further details of how researchers can develop ideas about the suitable range of threshold probabilities [9]. When researchers decide to use DCA for a model used for patient counseling, decision curves might be plotted for wider ranges of risk thresholds, even for the full range between 0 and 1.

Not using DCA to choose a risk threshold - We cannot use DCA to choose an optimal risk threshold. NB depends on the adopted risk threshold, not the other way around. More generally, the choice of a clinically appropriate threshold should not depend on the results of a study of a prediction model [16].

Reporting the test tradeoffs where appropriate - An increase in NB may not be worth the additional cost of using the best model. Investigators should consider reporting the test tradeoff, in particular when there are significant harms or costs associated with obtaining data for the model. When comparing two models, we can express the test tradeoff in terms of true positives and true negatives. We recommend reporting both.

CONCLUSION

DCA is a statistical method to evaluate whether a model has utility in supporting clinical decisions, and which of two models leads to the best decisions. It is therefore an essential validation tool on top of measures such as discrimination and calibration.

REFERENCES

1. E.W. Steyerberg Clinical prediction models: a practical approach to development, validation, and updating Springer, New York, NY (2009)
2. S.F. Shariat, M.W. Kattan, A.J. Vickers, P.I. Karakiewicz, P.T. Scardino Critical review of prostate cancer predictive tools Future Oncol, 5 (2009), pp. 1555-1584
3. A.J. Vickers, A.M. Cronin. Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). Urology, 76 (2010), pp. 1298-1301
4. B. Van Calster, D. Nieboer, Y. Vergouwe, B. De Cock, M.J. Pencina, E.W. Steyerberg. A calibration hierarchy for risk models was defined: from utopia to empirical data J Clin Epidemiol, 74 (2016), pp. 167-176
5. B. Van Calster, A.J. Vickers. Calibration of risk prediction models: impact on decision-analytic performance Med Decis Making, 35 (2015), pp. 162-169
6. N. Olchanski, J.T. Cohen, P.J. Neumann, J.B. Wong, D.M. Kent. Understanding the value of individualized information: the impact of poor calibration or discrimination in outcome prediction models Med Decis Making, 37 (2017), pp. 790-801
7. A.J. Vickers. Incorporating clinical considerations into statistical analyses of markers: a quiet revolution in how we think about data Clin Chem, 62 (2016), pp. 671-672
8. A.J. Vickers, E.B. Elkin. Decision curve analysis: a novel method for evaluating prediction models Med Decis Making, 26 (2006), pp. 565-574
9. A.J. Vickers, B. Van Calster, E.W. Steyerberg. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests BMJ, 352 (2016), p. i6
10. A.J. Vickers Prediction models: revolutionary in principle, but do they do more good than harm? J Clin Oncol, 29 (2011), pp. 2951-2952
11. A.R. Localio, S. Goodman. Beyond the usual prediction accuracy metrics: reporting results for clinical decision-making Ann Intern Med, 157 (2012), pp. 294-295
12. L. Holmberg, A.J. Vickers. Evaluation of prediction models for decision-making: beyond calibration and discrimination. PLoS Med, 10 (2013), p. e1001491
13. M.S. Pepe, H. Janes, C.I. Li. Net risk reclassification p values: valid or misleading?J Natl Cancer Inst, 106 (2014), p. dju041
14. V.P. Balachandran, M. Gonen, J.J. Smith, R.P. DeMatteo. Nomograms in oncology: more than meets the eye Lancet Oncol, 16 (2015), pp. e173-e180
15. M. Fitzgerald, B.R. Saville, R.J. Lewis. Decision curve analysis. JAMA, 313 (2015), pp. 409-410
16. K.F. Kerr, M.D. Brown, K. Zhu, H. Janes Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. J Clin Oncol, 34 (2016), pp. 2534-2540
17. E.W. Steyerberg, A.J. Vickers, N.R. Cook, et al. Assessing the performance of prediction models: a framework for traditional and novel measures Epidemiology, 21 (2010), pp. 128-138
18. K.G.M. Moons, D.G. Altman, J.B. Reitsma, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med, 162 (2015), pp. W1-W73
19. F.H. Schröder, J. Hugosson, M.J. Roobol, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up Lancet, 384 (2014), pp. 2027-2035

20. L. Van Neste, R.J. Hendriks, S. Dijkstra, et al. Detection of high-grade prostate cancer using a urinary molecular biomarker-based risk score Eur Urol, 70 (2016), pp. 740-748

21. N. Mottet, J. Bellmunt, M. Bolla, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol, 71 (2017), pp. 618-629

22. M.J. Roobol, E.W. Steyerberg, R. Kranse, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer Eur Urol, 57 (2010), pp. 79-85

23. N.J. Perkins, E.F. Schisterman. The inconsistency of "optimal" cut points obtained using two criteria based on the receiver operating characteristic curve Am J Epidemiol, 163 (2006), pp. 670-675

24. S.G. Pauker, J.P. Kassirer Therapeutic decision making: a cost-benefit analysis. New Engl J Med, 293 (1975), pp. 229-234

25. M.A. Hlatky, P. Greenland, D.K. Arnett, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association Circulation, 119 (2009), pp. 2408-2416

26. B. Van Calster, A.J. Vickers, M.J. Pencina, S.G. Baker, D. Timmerman, E.W. Steyerberg. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures Med Decis Making, 33 (2013), pp. 490-501

27. S.G. Baker, N.R. Cook, A.J. Vickers, B.S. Kramer Using relative utility curves to evaluate risk prediction J R Stat Soc A, 172 (2009), pp. 729-748

28. M.J. Pencina, R.B. D'Agostino Sr., E.W. Steyerberg. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med, 30 (2011), pp. 11-21

29. M.G.M. Hunink, M.C. Weinstein, E. Wittenberg, et al. Decision making in health and medicine: integrating evidence and values (ed. 2), Cambridge University Press, Cambridge, UK (2014)

30. S.G. Baker, B. Van Calster, E.W. Steyerberg. Evaluating a new marker for risk prediction using the test tradeoff: an update Int J Biostat, 8 (2012), p. 1

31. S.G. Baker, E. Schuit, E.W. Steyerberg, et al. How to interpret a small increase in AUC with an additional risk prediction marker: decision analysis comes through Stat Med, 33 (2014), pp. 3946-3959

32. S.G. Baker, B.S. Krame.r Evaluating prognostic markers using relative utility curves and test tradeoffs J Clin Oncol, 33 (2015), pp. 2578-2580

33. V. Rousson, T. Zumbrunn. Decision curve analysis revisited: overall net benefit, relationships to ROC curve analysis, and application to case-control studies BMC Med Inform Decis Making, 11 (2011), p. 45

34. M.J. Roobol, H.A. van Vugt, S. Loeb, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators Eur Urol, 61 (2012), pp. 577-583

35. N. Pereira-Azevedo, I. Braga, J.F. Verbeek, et al. Prospective evaluation on the effect of interobserver variability of digital rectal examination on the performance of the Rotterdam Prostate Cancer Risk Calculator Int J Urol, 24 (2017), pp. 826-832

36. J.P. Radtke, M. Wiesenfarth, C. Kesch, et al. Combined clinical parameters and multiparametric magnetic resonance imaging for advanced risk modeling of prostate cancer-patient-tailored risk stratification can reduce unnecessary biopsies

Eur Urol, 72 (2017), pp. 888-896

SUPPLEMENTARY MATERIALS

Formulas and Variants of NB

NB is calculated as:

$$NB = \frac{TP - odds(T) \times FP}{N}$$

Where TP is the number of true positives, FP the number of false positives, and N the sample size. NB divides by N, and hence gives the proportion of net true positives.
At the 10% threshold, the baseline model has NB=0.0393 (Table 2). This means that there are 3.9 net detected high-grade prostate cancers per 100 patients. Several measures have been proposed that are closely related to NB, and that lead to identical conclusions (see Appendix 1).[1-4]

NB of treat all is $NB_{TreatAll} = P - odds(T) \times (1 - P)$

Where P is the disease prevalence.
The net proportion of true negatives (i.e. avoided false positives) equals

$$\frac{NB - NB_{TreatAll}}{odds(T)}$$

Variants of NB

NB depends on disease prevalence P. When prevalence is low (e.g. when predicting high-grade prostate cancer), there are less potential true positives but more potential false positives than when prevalence is high (e.g. when predicting prostate cancer irrespective of grade). The maximum value of NB equals the prevalence, which occurs when the risk threshold T is 0 (i.e. we do not want to miss any case with disease at any cost), or at any threshold T for which classification is perfect (no false positives or false negatives).

NB/P is the 'standardized NB'1, with a maximum value of 1. In the case study, the prevalence is 8.6%. Therefore, the baseline model has a standardized NB of 0.46 at the 10% threshold.
A related metric, Relative Utility (RU), presents NB as the proportion of the maximum possible increase versus default:[3]

When NB of a model equals the NB of the best default strategy, RU is 0. Maximum NB equals prevalence, and then RU is 1. If NB of the model is worse than the NB of a default strategy, RU is negative. When T>P, RU equals standardized NB because max(NBTreatAll, NBTreatNone)=0. For the case study, the 10% risk threshold is higher than the prevalence,

$$RU = \frac{NB - max(NB_{TreatAll}, NB_{TreatNone})}{P - max(NB_{TreatAll}, NB_{TreatNone})}$$

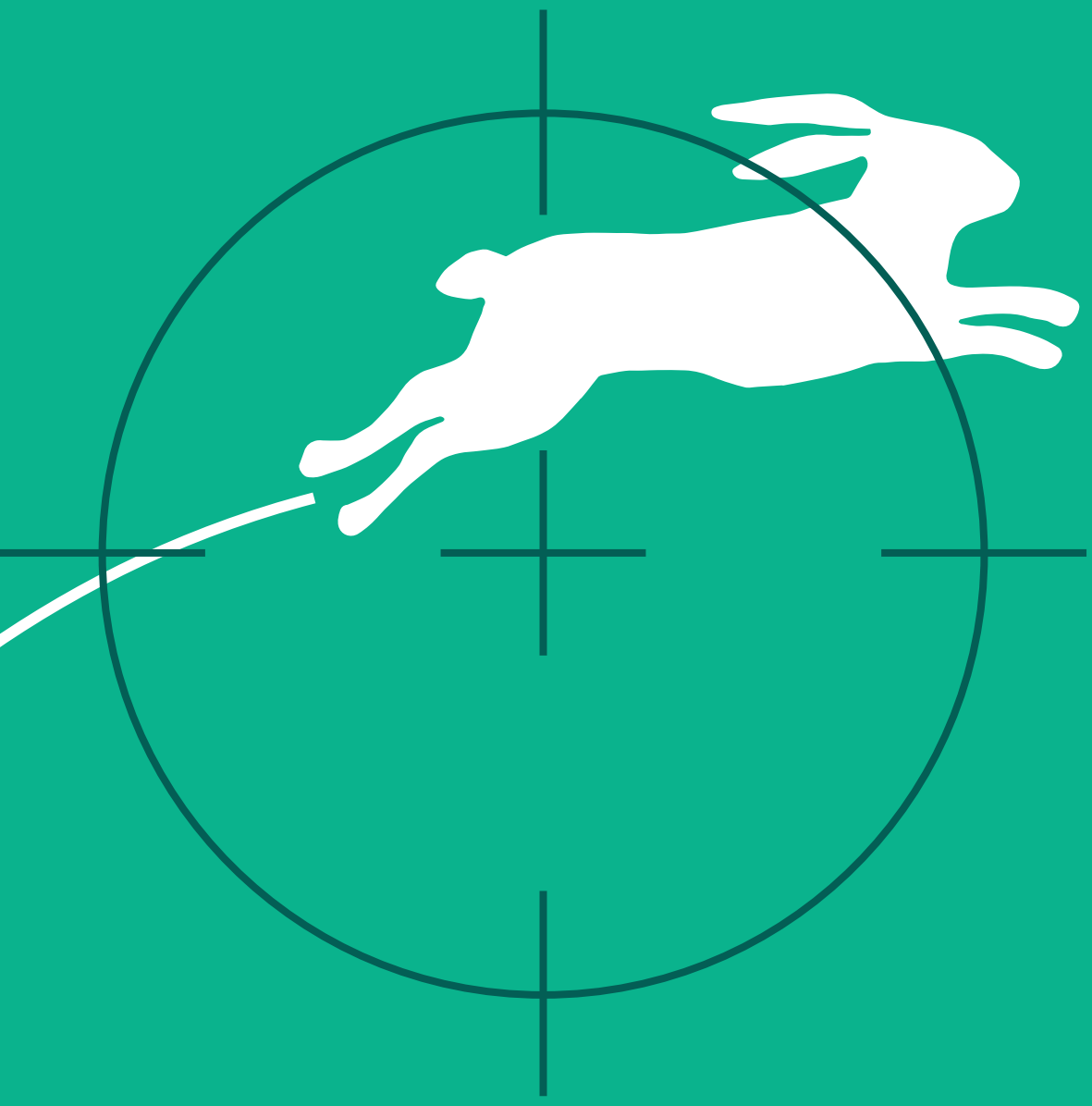
hence RU of the baseline model equals the standardized NB. Although it may appear as if standardized Net Benefit and RU are independent of prevalence, this is not the case.[5]
For model comparison, a weighted version of the Net Reclassification Improvement has been introduced (wNRI).[4] This can be rewritten as ΔNB/T, with ΔNB the difference in NB between the models.2
NB and its relative interpretations are entirely consistent, leading to identical conclusions.2 We prefer to work with NB, in order to describe utility on the original (absolute) scale.

References

[1] K.F. Kerr, M.D. Brown, K. Zhu and H. Janes, Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use, J Clin Oncol 34, 2016, 2534–2540.
[2] B. Van Calster, A.J. Vickers, M.J. Pencina, S.G. Baker, D. Timmerman and E.W. Steyerberg, Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures, Med Decis Making 33, 2013, 490–501.
[3] S.G. Baker, N.R. Cook, A.J. Vickers and B.S. Kramer, Using relative utility curves to evaluate risk prediction, J R Stat Soc A 172, 2009, 729–748.
[4] M.J. Pencina, R.B. D'Agostino, Sr. and E.W. Steyerberg, Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers, Stat Med 30, 2011, 11–21.
[5] J. Hilden, Prevalence-free utility-respecting summary indices of diagnostic power do not exist, Stat Med 19, 2000, 431–440.

Funding/Support and role of the sponsor:

B.V.C, L.W., J.Y.V., and E.C. are supported by the Research Foundation–Flanders (FWO) project G0B4716N, and Internal Funds KU Leuven (project C24/15/037). J.F.M V is supported by a grant of Prostate CancerUK and the SWOP (the Rotterdam Prostate Cancer Research Foundation). A.J.V. is supported by funds from the Sidney Kimmel Center for Prostate and Urologic Cancers, P50-CA92629 SPORE grant from the National Cancer Institute to Dr. H. Scher, the P30-CA008748 NIH/NCI Cancer Center Support Grant to MSKCC, and R01 CA179115 to Dr. A. Vickers. E.W.S. is supported by Patient-Centered Outcomes Research Institute (PCORI) Award (ME-1606-35555) and by the National Institutes of Health (U01NS086294). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI), its Board of Governors, or Methodology Committee. None of the funders had a role in the study design, data collection, data analysis, data interpretation, or the writing of the manuscript.



CHAPTER 5

Improving the Rotterdam European randomized study of screening for prostate cancer risk calculator for initial prostate biopsy by incorporating the 2014 international society of urological pathology Gleason grading and cribriform growth

Roobol MJ*
Verbeek JFM*
Van der Kwast T
Kümmerlin IP
Kweldam CF
Van Leenders GJLH

* These authors are joint first authors

Eur Urol. 2017 Jul;72(1):45-51

ABSTRACT

Background

The survival rate for men with International Society of Urological Pathology (ISUP) grade 2 prostate cancer (PCa) without invasive cribriform (CR) and intraductal carcinoma (IDC) is similar to that for ISUP grade 1. If updated into the European Randomized Study of Screening for Prostate Cancer (ERSPC Rotterdam) risk calculator number 3 (RC3), this may further improve upfront selection of men who need a biopsy.

Objective

To improve the number of possible biopsies avoided, while limiting undiagnosed clinically important PCa by applying the updated RC3 for risk-based patient selection.

Design, setting, and participants

The RC3 is based on the first screening round of the ERSPC Rotterdam, which involved 3616 men. In 2015, histopathologic slides for PCa cases (n=885) were re-evaluated. Low-risk (LR) PCa was defined as ISUP grade 1 or 2 without CR/IDC. High-risk (HR) PCa was defined as ISUP grade 2 with CR/IDC and PCa with ISUP grade≥3.

Outcome measurements and statistical analysis

We updated the RC3 using multinomial logistic regression analysis, including data on age, PSA, DRE, and prostate volume, for predicting LR and HR PCa. Predictive accuracy was quantified using receiver operating characteristic analysis and decision curve analysis.

Results and limitations

Men without PCa could effectively be distinguished from men with LR PCa and HR PCa (area under the curve 0.70, 95% confidence interval [CI] 0.68-0.72 and 0.92, 95% CI 0.90-0.94). At a 1% risk threshold, the updated calculator would lead to a 34% reduction in unnecessary biopsies, while only 2% of HR PCa cases would be undiagnosed.

Conclusions

A relatively simple risk stratification tool augmented with a highly sensitive contemporary pathologic biopsy classification would result in a considerable decrease in unnecessary prostate biopsies and overdiagnosis of potentially indolent disease.

Patient summary

We improved a well-known prostate risk calculator with a new pathology classification system that better reflects disease burden. This new risk calculator allows individualized prediction of the chance of having (potentially aggressive) biopsy-detectable prostate cancer and can guide shared decision-making when considering prostate biopsy.

INTRODUCTION

Although there is level A evidence that prostate-specific antigen (PSA)-based screening for prostate cancer (PCa) reduces metastatic disease and PCa death, there are a number of potential harms, including unnecessary biopsies and overdiagnosis with subsequent overtreatment of PCa considered as low-risk disease, often defined as Gleason score (GS) 3 + 3 (International Society of Urological Pathology [ISUP] grade 1) PCa [1-3]. The balance of benefits and harms can be positively influenced by replacing the purely PSA-based strategy with multivariate assessed, risk-based selection criteria [4-6]. Earlier studies showed that risk calculator number 3 (RC3) of the European Randomized Study of Screening for Prostate Cancer (ERSPC; www.erspc.org) based on the Rotterdam cohort is an adequate tool for risk stratification of men before prostate biopsy [7-9]. The RC3 uses prebiopsy information such as PSA, digital rectal examination (DRE) outcome, and prostate volume to predict the probability of a biopsy-detectable PCa and, more specifically, a potentially aggressive PCa defined as GS ≥ 3 + 4 (ISUP grade ≥2).

Recent studies suggest that survival rates for GS 3 + 4 (ISUP grade 2) PCa in the absence of an invasive cribriform growth pattern and intraductal carcinoma (CR−/IDC−) is similar to that for ISUP grade 1 PCa and that these men carry a substantial risk of being overdiagnosed and overtreated [10-12]. Incorporation of this additional tumor-specific information into the risk stratification could further improve upfront selection of men who need or currently do not need a prostate biopsy and, if applicable, would be of aid in treatment decisions.

The aim of this study was to further augment the RC3 with information from a multivariable multinominal regression to predict the chance of biopsy-detectable PCa defined as low- or high-risk disease on the basis of the ISUP Gleason grading system and invasive CR and/or IDC tumor growth patterns.

PATIENTS AND METHODS

Study Population

The ERSPC Rotterdam RC3 (available in the app store and on www.prostatecancer-riskcalculator.com) is suitable for men facing their first prostate biopsy. The RC3 is based on 3616 men biopsied in the first round of screening in the Rotterdam ERSPC section (age 54–74 yr; lateral sextant biopsy was indicated for PSA ≥3.0; inclusion period November 1993–March 2000) [6]. A total of 885 PCa cases were detected. Initial Gleason grading was performed by one pathologist (TvdK) following the original Gleason grading system [13].

Pathology evaluation

In 2015, all histopathology slides still available were re-evaluated by one urogenital pathologist and two pathologists in training (GvL, CK, and IK), who were blinded to patient information and biopsy outcome [10]. From the 885 PCa cases in the original ERSPC cohort, 36 (4.0%) biopsy tissue slides were no longer available for pathologic re-evaluation. Those cases were removed from the analysis. No other values were missing. For each biopsy core, the Gleason score, the presence of IDC, and the presence of Gleason grade 4 and 5 growth patterns were recorded [10-12]. Gleason grading was performed according to the 2014 ISUP recommendations [14]. The label CR/IDC+ was applied to patients who had invasive CR carcinoma or/and IDC; otherwise the label CR−/IDC− was applied. We defined a three-category outcome variable a priori: no PCa and two categories of PCa based on the ISUP 2014 Gleason grading and CR/IDC status. Low-risk (LR) PCa was defined as ISUP grade 1 or 2 and CR−/IDC−. High-risk (HR) was defined as ISUP grade 2 PCa with CR/IDC+ and all PCa with ISUP grade ≥3.

Statistical analysis

We used multivariable multinominal logistic regression modeling to estimate the risk of no PCa, LR PCa, and HR PCa. Similar to the original RC3, we included the predictors PSA (2log transformed and centered), DRE outcome (abnormal/normal, coded as 1/0, respectively), transrectal ultrasound (TRUS) outcome, and TRUS-assessed prostate volume. In addition, we developed a second model without TRUS outcome but including three categories of DRE-assessed prostate volume based on TRUS measurements. TRUS-assessed prostate volume was recoded as follows: <30 cm3 as 25 cm3; between 30 and 50 cm3 as 40 cm3; and ≥50 cm3 as 60 cm3, in accordance with a previously described study [15]. An additional candidate predictor was age (in years).

The predictive accuracy was quantified using the area under the curve (AUC) for the receiver operator characteristic (ROC) analysis [16]. Bootstrap resampling was used for internal validation. Comparison of ROC curves was performed using the method of DeLong. The DRE-based model, a model that does not require TRUS, is judged to be relevant, especially for implementation in primary care. The predictive abilities of the original DRE-based model and the improved DRE-based model were compared. Since the outcome of the two models is different, we combined the LR and HR PCa probabilities of the improved DRE-based model and compared this combined probability with the probability of any PCa using the original DRE-based model. In addition, the predictive ability for HR PCa was assessed for both models; it should be noted that HR is defined differently in the two models, and LR PCa patients were not separated from patients without PCa in the original DRE-based model. The clinical impact of the risk prediction model was analyzed using decision curve analysis and was focused on the DRE-based model; as a visual comparison, the net benefit curve for a

PSA-only model is displayed. The clinical importance of the model can be represented by the net benefit ratio, which weighs the benefits (detecting cancer) versus the harms (unnecessary biopsy) [17], [18]. Standard statistical software was used (SPSS version 24.0, IBM Corp, Armonk, NY, USA and R version 3.3.1, R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Participants

Prostate cancer was found in 849 (24%) biopsy slides. Table 1 shows the original and the new ISUP 2014 Gleason grading including CR/IDC status. Among men with previously classified GS 3 + 3 biopsy outcome, a total of 50 (9%) biopsy specimens were graded as HR PCa after re-evaluation. For men with an initial GS 3 + 4 biopsy outcome, 91 (42%) were re-classified as LR PCa. Eleven (14%) men initially graded as GS > 3 + 4 were reclassified as LR PCa.

Table 1. Cross-comparison of the original Gleason score classified as ≤3 + 3, 3 + 4, and >3 + 4 versus the 2014 ISUP Gleason grading.

Original score	Re-evaluation using the proposed scheme, *n* (%)		Total
	ISUP grade 1 or 2 and CR^-/IDC^-	ISUP grade 2 and CR^+/IDC^+ or ISUP grade >2	
Gleason ≤ 3 + 3	505 (91.0)	50 (9.0)	555
Gleason 3 + 4	91 (42.3)	124 (57.7)	215
Gleason > 3 + 4	11 (13.9)	68 (86.1)	79

ISUP = International Society of Urological Pathology Gleason; CR = cribriform growth pattern; IDC = intraductal carcinoma.

Of all PCa cases, after applying the ISUP 2014 grading and CR/IDC status, 607 men were graded as LR and 242 as HR, leaving 2731 men with no PCa detected in their first screening round. Looking at differences between no PCa, LR PCa, and HR PCa, a rising PSA, decreasing prostate volume, older age, and increasing number of abnormal DRE findings were seen (Table 2).

Table 2. Patient characteristics stratified according to the 2014 International Society of Urological Pathology Gleason grading and the presence of a cribriform growth pattern (CR) or intraductal carcinoma (IDC).

Characteristics	No PCa	Low-risk PCa	High-risk PCa
Patients (*n*)	2731	607	242
Median PSA, ng/ml (IQR)	4.0 (2.5–5.7)	5.0 (3.6–7.3)	8.0 (5.1–15.3)
Median TRUS volume, ml (IQR)	42.7 (32.7–56.5)	37.3 (29.1–49.1)	36.5 (28.6–45.7)
Median age, yr (IQR)	65.6 (61.0–69.7)	65.8 (61.6–69.6)	67.8 (64.0–71.6)
Abnormal DRE, *n* (%)	836 (30.6)	243 (40.0)	175 (72.3)
Abnormal TRUS, *n* (%)	795 (29.1)	233 (38.4)	181 (74.8)
Gleason score, *n* (%)			
≤6	–	392 (64.6)	–
3 + 4	–	215 (35.4)	47 (19.4)
4 + 3	–	–	87 (35.9)
≥4 + 4	–	–	108 (44.7)
CR/IDC presence, *n* (%)	–	–	163 (67.4)

PCa = prostate cancer; DRE = digital rectal examination; TRUS = transrectal ultrasound.

Development of the multinomial model

In addition to the predictors for biopsy outcome in the original TRUS-based risk calculator, age was also a significant predictor for HR PCa (odds ratio 1.06, 95% confidence interval [CI] 1.03–1.09; Table 3). As expected from Table 2, large prostate volume for similar age, PSA, and DRE and TRUS outcomes reduced the risk of having both LR and HR PCa. The cancer detection rate (number of cancers found among eligible men) was 17% (607/3580) for LR PCa and 7% (242/3580) for HR PCa. Men without PCa could be well distinguished from men with LR PCa in the TRUS-based model (AUC 0.73, 95% CI 0.70–0.75) and DRE-based model (AUC 0.70, 95% CI: 0.68–0.72). The AUCs for no PCa versus HR PCa were 0.94 (95% CI 0.92–0.95), and 0.91 (95% CI 0.89–0.93), respectively (Table 3). The TRUS-based model yielded no higher discrimination than the DRE-based model for LR PCa prediction ($p = 0.09$) and HR PCa prediction ($p = 0.07$). The predictive ability of the original DRE-based RC3 (AUC 0.79, 95% CI 0.77–0.81) was comparable to that of the improved DRE-based RC3 (AUC 0.77, 95% CI 0.75–0.79) in predicting any PCa ($p = 0.09$). For HR PCa the AUC was significantly higher in the improved DRE-based RC3 (AUC 0.91, 95% CI 0.89–0.93) compared to the original RC3 (AUC 0.84, 95% CI 0.82–0.86; $p < 0.001$). Considering the broader clinical applicability of the DRE-based model and the lack of superior discriminative ability of the TRUS-based model, the decision curve analysis was continued with the DRE-based model only.

Table 3. Results for multivariable multinomial logistic regression analysis using participants without PCa as the reference group.

Model and predictor	OR (95% CI)	
	Low-risk PCa	High-risk PCa
DRE model		
PSA	2.29 (2.06–2.55)	5.05 (4.27–5.98)
Age	1.00 (0.99–1.03)	1.06 (1.03–1.10)
Abnormal DRE	1.91 (1.56–2.33)	7.50 (5.38–10.46)
Prostate volume (class via DRE)	0.26 (0.20–0.32)	0.11 (0.07–0.16)
Area under the curve (95% CI)	0.70 (0.68–0.72)	0.91 (0.89–0.93)
TRUS model		
PSA	2.60 (2.33–2.91)	5.78 (4.65–6.69)
Age	1.02 (1.00–1.04)	1.08 (1.05–1.12)
Abnormal DRE	1.74 (1.42–2.13)	4.20 (2.95–6.00)
Prostate volume (continuous via TRUS)	0.30 (0.25–0.36)	0.16 (0.12–0.22)
Abnormal TRUS	1.87 (1.53–2.23)	7.18 (4.99–10.33)
Area under the curve (95% CI)	0.73 (0.70–0.75)	0.94 (0.92–0.95)

PCa = prostate cancer; OR = odds ratio; CI = confidence interval; DRE = digital rectal examination; PSA = prostate-specific antigen; TRUS = transrectal ultrasound.

Decision curve analysis

Application of different risk thresholds to the newly updated DRE-based risk model resulted in decreases in the numbers of biopsies and LR/HR PCas detected or missed (Table 4 and Supplementary Table 1). A risk threshold of 1% for a positive HR PCa biopsy would result in 34% fewer biopsies (n = 1207), with a high true-positive rate of 87% (736/849). With this strategy, 13% of all PCas would have been missed (n = 113). Of these PC cases, the majority (96%, n = 108) would have been classified as LR PCa. Five clinically important PCas would have been missed (false-negative rate 5/242, 2%). The overall performance of the model can be assessed by weighing the number of cancers detected against the number of unnecessary biopsies over a range of thresholds, as displayed in a decision curve and compared to a PSA-only model in Supplementary Figure 1. The updated RC3 and original RC3 led to a similar reduction in unnecessary biopsies (34% vs 33%) and missed diagnosis of any PCa (13% vs 14%). However, the newly updated DRE model missed only 2% of HR PCa, compared to 7% with the original risk calculator [6]. Hence, at a fixed sensitivity of 93% for HR PCa, the original risk calculator achieved a 33% biopsy reduction, whereas the new model could avoid 61% of potentially unnecessary biopsies.

Table 4. Number of biopsies (saved) resulting in LR and HR PCs detected per missed numbers at initial screening with various cutoffs for detecting high-risk PC with the DRE-based risk calculator.

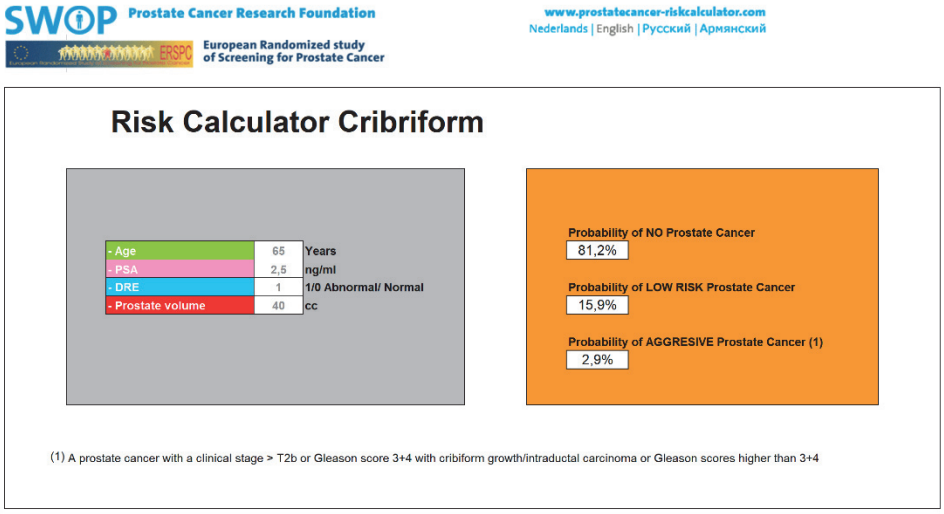
Risk	(A) No. of men biopsied	(B) No. of biopsies saved (% of total A)	(C) No. of PCs detected (% PPV)	D: No. of LR PCs detected (% of A)	(E) No. of LR PCs missed (% of total D)	(F) No. of HR PCs detected (% of A)	(G) No. of HR PCs missed (% of total F)
Total	3580	0	849 (23.7)	607 (17.0)	0	242 (6.8)	0
1%	2373	1207 (33.7)	736 (31.0)	499 (21.0)	108 (17.8)	237 (10.0)	5 (2.1)
1.5%	2017	1563 (43.7)	675 (33.5)	443 (22.0)	164 (27.0)	232 (11.5)	10 (4.1)
2%	1749	1831 (51.1)	635 (36.3)	408 (23.3)	199 (32.8)	227 (13.0)	15 (6.2)
3%	1412	2168 (60.6)	581 (41.1)	357 (25.3)	250 (41.2)	224 (15.9)	18 (7.4)
5%	1024	2556 (71.4)	493 (48.1)	284 (27.7)	323 (53.2)	209 (20.4)	33 (13.6)

LR = low risk; HR = high risk; PC = prostate cancer; DRE = digital rectal examination; PPV = positive predictive value.

Example of clinical application

To illustrate the updated DRE-based model, Fig. 1A shows the prediction for a 65-yr-old man with PSA 3.5 ng/ml, DRE-assessed prostate volume of 40 cm3, and normal DRE findings. The risk of LR PCa and HR PCa is 13% and 1%, respectively. An abnormal DRE outcome (Fig. 1B) would increase his risk of HR PCa to 6%. For this patient, with a predicted 1% probability of developing a clinically important PCa, the risk calculator may aid the patient and urologist in deciding whether to refrain from biopsy. An important consideration is that undergoing a biopsy is uncomfortable and has a risk of approximately 3% for complications such as infection and sepsis [3]. Another argument is that there is a high chance (13%) of finding indolent PCa (potential overdiagnosis). For this hypothetical patient with relatively low PSA, abnormal DRE findings would result in a risk of almost 6% for HR PCa. Therefore, for thresholds ranging from 1% to 5% for HR PCa, biopsy is recommended.

(A)



(B)

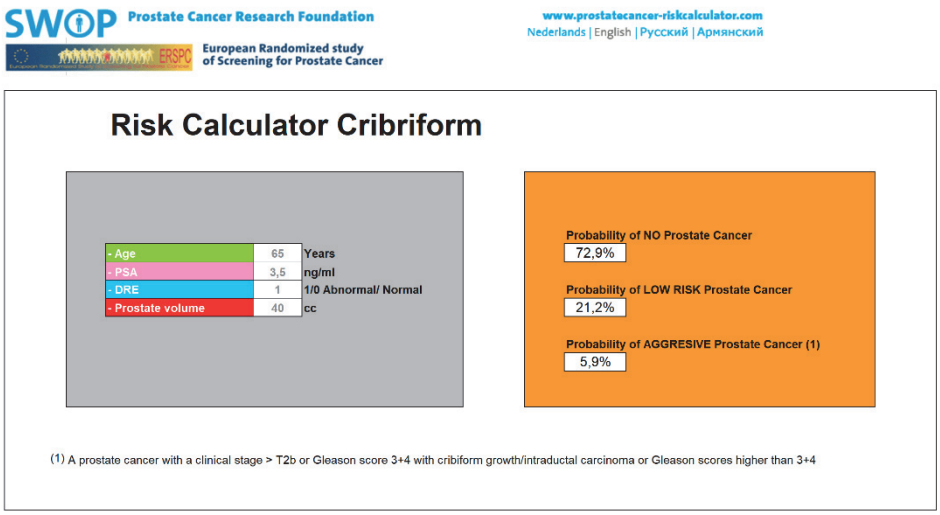


Figure 1. Risk for (A) a 65-yr-old man with prostate-specific antigen (PSA) of 3.5 ng/ml, a normal digital rectal examination (DRE), and a prostate volume of 40 cm3 and (B) a 65-yr-old man with PSA of 3.5 ng/ml, abnormal DRE, and prostate volume of 40 cm3.

DISCUSSION

The newly updated risk calculator based on the 2014 ISUP grading system and CR/IDC status is an accurate tool for predicting the risk of (clinically important) PCa. It facilitates shared decision-making when deciding on whether to perform a biopsy, a procedure that has some side effects [3], [19]. This new risk calculator, like the original TRUS- and DRE-based ones, shows high discriminatory value for predicting HR PCa [7]. CR/IDC morphology in addition to Gleason grading could provide a better understanding of disease burden. This is important because guidelines for active surveillance (AS), which are based on several large ongoing cohorts, consider some men with GS 3 + 4 disease eligible for AS [20]. Kweldam et al [10] reported 15-yr disease-specific survival probabilities among men with different GS and CR/IDC status in biopsy specimens. The survival outcome did not differ between patients with GS 3 + 4 without the presence of CR/IDC and men with GS 6 [10]. This finding could imply that men with biopsy GS 3 + 4 and CR−/IDC− PCa could also be candidates for AS.

A large AS cohort with long-term follow-up for men with GS 3 + 4 PCa showed that patients who developed metastases more often had GS 3 + 4 disease at the time of diagnosis [21]. Knowing that more than half of the men with GS 3 + 4 PCa in our study cohort had in fact CR/IDC+ status, it is possible that patients in the Klotz cohort had similar characteristics that could explain the rate of metastases observed. In the recent published ProtecT study, 1643 patients with localized PCa were randomized to different treatment options (active monitoring, radical prostatectomy, external radiation) [22]. After median follow-up of 10 yr, no differences in PCa-specific deaths were found. However, a significantly higher number of men in the active monitoring arm suffered metastatic disease. These findings should be interpreted with caution. First, current AS protocols have more restrictive entry criteria than ProtecT had in the active monitoring arm (eg, randomizing GS ≥8 to this arm). Furthermore, no scheduled re-biopsy was provided in the ProtecT active monitoring follow-up protocol. Both arguments could contribute to the significantly higher number of men in the AS arm who suffered metastatic disease. Further research is needed to confirm whether the prognostic value of CR/IDC status is similar in contemporary protocol-based AS protocols.

To the best of our knowledge, this is the first study incorporating the new ISUP 2014 Gleason grading and including CR/IDC status in the existing risk calculator for predicting biopsy outcome for both LR and HR PCa. Although it has been shown that systematic PSA-based screening reduces PCa-specific mortality [23], [24], population-based PCa screening programs are not yet acceptable because of the high numbers of unnecessary tests and the detection of indolent PCa that would never cause any harm (overdiagnosis) [25]. Risk calculators improve the predictive accuracy of PSA testing in detecting PCa. The diagnostic accuracy of our updated DRE model is higher than the PCPT for detecting high-grade PCa

(AUC 0.71, 95% CI 0.67–0.75) [9]. To study the improvement in diagnostic accuracy, we compared our updated model with the original DRE-based model. The updated RC3-based strategy would lead to a further reduction in undiagnosed clinically important PCa, while achieving the same reduction in unnecessary biopsies. Incorporation of the 2014 ISUP Gleason grading and CR/IDC status gives a better understanding of clinically important PCa, which we would not want to miss. The proposed low cutoff of 1% risk circumvents this while keeping the benefit of avoiding unnecessary biopsies to a similar level as for our original RC (ie, ~33%).

In daily practice, the updated DRE-based risk calculator is user-friendly owing to its easily retrievable predictor information. There is the possibility that general practitioners will work with this risk calculator, since it bypasses the need for referral or expensive and time-consuming imaging examinations such as TRUS and magnetic resonance imaging. Although prostate volume estimated by TRUS provides better discrimination and calibration in risk assessment, the easy-to-use DRE-based model shows similar performance.

This study has some limitations. As no external validation has yet been performed—only internal validation with bootstrap techniques—the discriminative power could be lower. Another limitation is that the model is based on the original biopsy strategy using TRUS-guided systematic sextant biopsy. This could have resulted in underestimation of the actual cancer risk. However, the original RC3 has been validated many times in contemporary clinical settings with good results [26, 27]. DRE-assessed volume was not assessed by the urologist but calculated using the TRUS data, which is a potential limitation. A recent study on the external validation of the original DRE-based RC3, however, showed good performance and confirmed the earlier reported good correlation between TRUS prostate volume and DRE-assessed prostate volume [7, 28]. Finally, men without CR/IDC growth but high-volume disease were categorized as LR PCa in this study, which might be counterintuitive. However, we have shown that GS 3 + 4 CR−/IDC− PCa, including high-volume disease, is similar to GS 6 disease with respect to, for example, survival. Furthermore, tumor volume is strongly associated with the presence of CR/IDC. Patients with GS 3 + 4 CR/IDC+ had median biopsy tumor involvement of 40%, in contrast to 23% among men lacking this pattern [10, 12].

In our study approximately 10% of the patients initially considered to have LR PCa were upgraded to HR PCa. Conversely, 33% ([91 + 13]/312) were downgraded. It is likely that some patients defined as LR in our study could later on be proven to have HR PCa owing to the possibility of biopsy sample error. However, this would probably be mitigated by the fact that most men with LR PCa often undergo an MRI scan and/or targeted biopsy, especially when considering AS as the initial treatment option [29, 30].

CONCLUSIONS

We have improved the easy-to-use DRE-based ERSPC risk calculator for predicting (clinically important) PCa by incorporating a highly sensitive contemporary pathologic biopsy classification of findings suitable for men at initial biopsy. If implemented into daily clinical practice, this could considerably decrease the number of unnecessary prostate biopsies and overdiagnosis of potentially indolent disease.

REFERENCES

1. F.H. Schröder, J. Hugosson, M.J. Roobol, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet, 384 (2014), pp. 2027-2035
2. C. Buzzoni, A. Auvinen, M.J. Roobol, et al. Metastatic prostate cancer incidence and prostate-specific antigen testing: new insights from the European Randomized Study of Screening for Prostate Cancer. Eur Urol, 68 (2015), pp. 885-890
3. S. Loeb, A. Vellekoop, H.U. Ahmed, et al. Systematic review of complications of prostate biopsy. Eur Urol, 64 (2013), pp. 876-892
4. M.J. Roobol, S.V. Carlsson Risk stratification in prostate cancer screening. Nat Rev Urol, 10 (2013), pp. 38-48
5. H. Gronberg, J. Adolfsson, M. Aly, et al. Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. Lancet Oncol, 16 (2015), pp. 1667-1676
6. M.J. Roobol, E.W. Steyerberg, R. Kranse, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. Eur Urol, 57 (2010), pp. 79-85
7. M.J. Roobol, H.A. van Vugt, S. Loeb, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators. Eur Urol, 61 (2012), pp. 577-583
8. R. Kranse, M. Roobol, F.H. Schröder A graphical device to represent the outcomes of a logistic regression analysis. Prostate, 68 (2008), pp. 1674-1680
9. K.S. Louie, A. Seigneurin, P. Cathcart, P. Sasieni. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. Ann Oncol, 26 (2015), pp. 848-864
10. C.F. Kweldam, I.P. Kummerlin, D. Nieboer, et al. Disease-specific survival of patients with invasive cribriform and intraductal prostate cancer at diagnostic biopsy. Mod Pathol, 29 (2016), 630-636
11. C.F. Kweldam, M.F. Wildhagen, E.W. Steyerberg, C.H. Bangma, T.H. van der Kwast, G.J. van Leenders. Cribriform growth is highly predictive for postoperative metastasis and disease-specific death in Gleason score 7 prostate cancer. Mod Pathol, 28 (2015), pp. 457-464
12. C.F. Kweldam, I.P. Kümmerlin, D. Nieboer, et al. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. Eur J Cancer, 66 (2016), 26-33
13. D.F. Gleason, G.T. Mellinger. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol, 111 (1974), pp. 58-64
14. J.I. Epstein, L. Egevad, M.B. Amin, et al. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. Am J Surg Pathol, 40 (2016), pp. 244-252
15. M.J. Roobol, F.H. Schröder, J. Hugosson, et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group. World J Urol, 30 (2012), pp. 149-155
16. E.W. Steyerberg, A.J. Vickers, N.R. Cook, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 21 (2010), pp. 128-138
17. A.J. Vickers, E.B. Elkin. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making, 26 (2006), pp. 565-574

18. K.F. Kerr, M.D. Brown, K. Zhu, H. Janes. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. J Clin Oncol, 34 (2016), pp. 2534-2540

19. S. van den Heuvel, S. Loeb, X. Zhu, et al. Complications of initial prostate biopsy in a European randomized screening trial. Am J Clin Exp Urol, 1 (2013), pp. 66-71

20. S.M. Bruinsma, C.H. Bangma, P.R. Carroll, et al. Active surveillance for prostate cancer: a narrative review of clinical guidelines. Nat Rev Urol, 13 (2016), pp. 151-167

21. L. Klotz, D. Vesprini, P. Sethukavalan, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol, 33 (2015), pp. 272-277

22. F.C. Hamdy, J.L. Donovan, J.A. Lane, et al. 10-Year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. N Engl J Med, 375 (2016), pp. 1415-1424

23. J. Hugosson, S. Carlsson, G. Aus, et al. Mortality results from the Goteborg randomized population-based prostate-cancer screening trial. Lancet Oncol, 11 (2010), pp. 725-732

24. F.H. Schröder, J. Hugosson, M.J. Roobol, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med, 366 (2012), pp. 981-990

25. D. Ilic, M.M. Neuberger, M. Djulbegovic, P. Dahm. Screening for prostate cancer. Cochrane Database Syst Rev, 2013 (2013), p. CD004720

26. H.A. van Vugt, M.J. Roobol, R. Kranse, et al. Prediction of prostate cancer in unscreened men: external validation of a risk calculator. Eur J Cancer, 47 (2011), pp. 903-909

27. Gayet M, Mannaerts CK, Nieboer D, et al. Prediction of Prostate Cancer: External Validation of the ERSPC Risk Calculator in a Contemporary Dutch Clinical Cohort. Eur Urol Focus.Mar;4(2):228-234.

28. N. Pereira-Azevedo, I. Isaac Braga, J.F.M. Verbeek, et al. A prospective evaluation on the effect of inter-observer variability of DRE on the performance of the DRE based Rotterdam Prostate Cancer Risk Calculator Accepted for the 2017 Annual Meeting of the American Urological Association Education and Research Inc., Boston, MA, USA. May 12–16 (2017)

29. A.R. Alberts, I.G. Schoots, L.P. Bokhorst, G.J. van Leenders, C.H. Bangma, M.J. Roobol Risk-based patient selection for magnetic resonance imaging-targeted prostate biopsy after negative transrectal ultrasound-guided random biopsy avoids unnecessary magnetic resonance imaging scans. Eur Urol, 69 (2016), pp. 1129-1134

30. I.G. Schoots, M.J. Roobol, D. Nieboer, C.H. Bangma, E.W. Steyerberg, M.G. Hunink. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur Urol, 68 (2015), pp. 438-450

SUPPLEMENTARY MATERIALS

Supplementary Table 1: Number of biopsies (saved) resulting in low risk and high-risk cancers detected per missed numbers at initial screening with various cut-offs for detecting low risk PCa (3%, 5%, 12.5%, 20%) with the DRE based risk calculator.

Risk	A	B	C	D	E	F	G
Total*	3850	0	849 (23.7)	607 (17.0)	0	242 (6.8)	0
3%	3518	62 (1.7)	848 (24.1)	607 (17.0)	0 (0)	241 (6.9)	1 (0.4)
5%	3381	199 (5.6)	842 (24.9)	603 (17.8)	4 (0.7)	239 (7.1)	3 (1.2)
12.5%	2244	1336 (37.3)	702 (31.3)	480 (21.4)	127 (20.9)	222 (10.0)	20 (8.3)
20%	1213	2367 (66.1)	529 (43.6)	344 (28.4)	263 (43.3)	185 (15.3)	57 (23.6)

* Biopsy all men. PPV: Positive Predictive Value, PCa= Prostate Cancer, LR= Low Risk, HR= High Risk

A: No. of men biopsied

B: No. biopsies saved (% of total A)

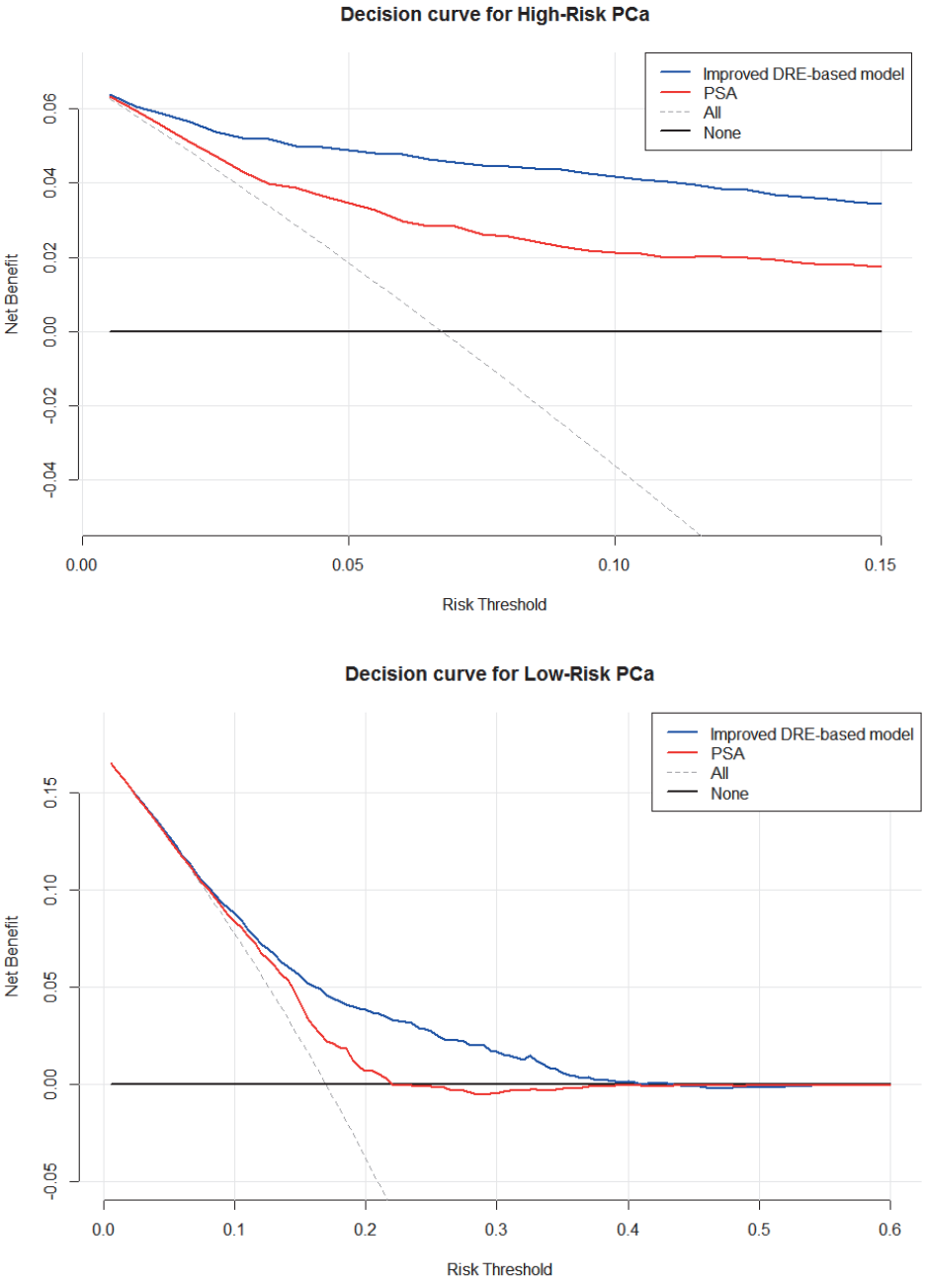
C: No. PC detected (% PPV)

D: No. of LR PCa detected (% of A)

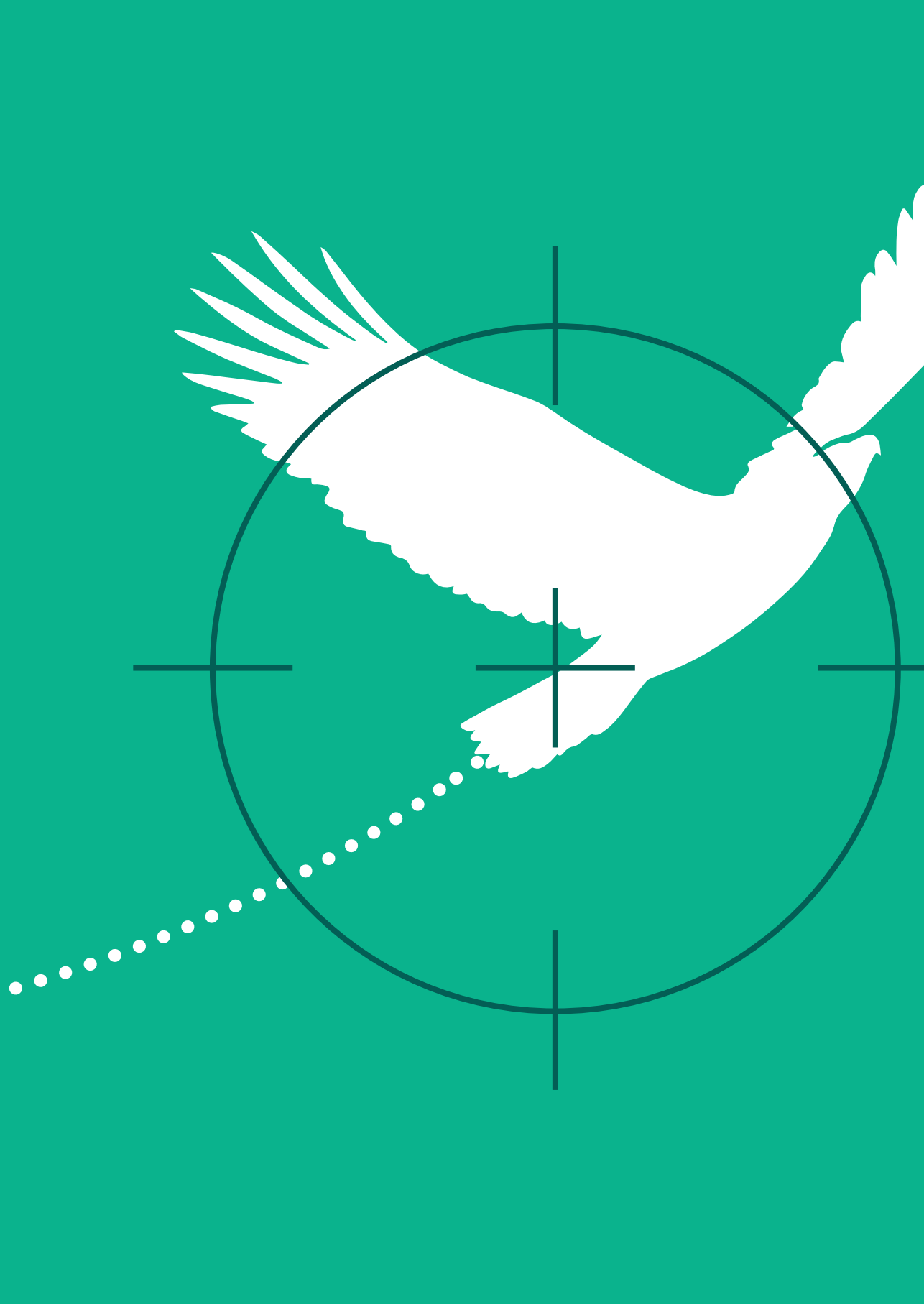
E: Missed LR PCa (% of total D)

F: No. of HR PCa detected (% of A)

G: Missed HR Pca (% of total F)



Supplementary Figure 1. Decision curves for the improved DRE-based model vs. only PSA-based calculator predicting high-risk PCa (above panel) and low-risk PCa (below panel) on men with an initial biopsy. Blue line: improved DRE based model. Red line: PSA-based model. Gray dotted line: assume all men receive biopsy. Black solid line at zero: assume no men receive a biopsy.



CHAPTER 6

Reducing unnecessary biopsies while detecting clinically significant prostate cancer including cribriform growth with the ERSPC Rotterdam risk calculator and 4Kscore

Verbeek JFM
Bangma CH
Kweldam CF
Van der Kwast T
Kümmerlin IP
Van Leenders GJLH
Roobol MJ

Urol Oncol. 2019 Feb;37(2):138-144

ABSTRACT

Introduction

The use of risk calculators predicting clinically significant prostate cancer (csCaP) on biopsy reduces unnecessary biopsies and overdiagnosis of indolent disease compared to a Prostate Specific Antigen (PSA) strategy. Updating these tools using more specific outcome measures and contemporary predictors could potentially lead to further reductions. Our objective was to assess clinical impact of the 4 kallikrein (4K) score, the Rotterdam Prostate Cancer Risk Calculator (RPCRC), and the combination of both for predicting csCaP based on the latest International Society of Urological Pathology grading system and cribriform growth pattern.

Materials and methods

Our prospective cohort consisted of 2,872 men from the first screening round in the European Randomized Study of Screening for Prostate Cancer Rotterdam; biopsy indication PSA ≥ 3.0. The predictive performance of the 4Kscore, RPCRC, and the combination of RPCRC with 4Kscore were assessed with area under the receiver operator characteristic curve (AUC) and calibration plots. Decision curve analysis was used to evaluate the reduction of unnecessary biopsy and indolent CaP.

Results

The csCaP was present in 242 (8%) men, and indolent CaP in 578 (20%). The 4Kscore and RPCRC had similar high AUCs (0.88 vs. 0.87; $P = 0.41$). The 4Kscore-RPCRC combination improved AUC to 0.89 compared to 4Kscore ($P < 0.01$) and RPCRC ($P < 0.01$). The RPCRC and 4Kscore reduced the number of Bx with 42 and 44, respectively, per 100 men at risk compared to a ≥PSA 3.0 strategy without increasing missed csCaP. The RPCRC-4Kscore combination resulted in a slight additional net reduction of 3.3 biopsies per 100 men.

Conclusions

The RPCRC and 4Kscore had similar reductions of unnecessary biopsies and overdiagnosis of indolent disease. Combination of both models slightly reduced unnecessary biopsies further. Gain in net benefit must, however, be weighed against additional costs and availability of tests.

INTRODUCTION

Prostate Specific Antigen (PSA)-based prostate cancer (CaP) screening is beneficial in terms of mortality reduction, however its main drawbacks are overdiagnosis and overtreatment of indolent CaP [1]. A more and more used strategy to limit overtreatment of indolent CaP cancer is the use of active surveillance as initial treatment [2]. To improve screening efforts further, a balance must be found between minimizing overdiagnosis, and optimizing the timely detection of potentially deadly disease [3]. Preferably only men with a clinically significant CaP (csCaP) should be identified and diagnosed. In addition, the therapeutic intervention of choice should have minimal impact on quality of life, to maximize the gain in quality adjusted life years.

For several decades, multivariable prediction tools have been constructed with the aim of selectively predicting the presence of csCaP on biopsy [4-6]. The Rotterdam Prostate Cancer Risk Calculator (RPCRC) is an example of these and is available as an app. It was developed on the basis of the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC). The RPCRC has demonstrated its value in predicting CaP risk and in avoiding unnecessary testing [4], [7, 8]. However, risk calculators are not perfect and with the development of new biomarkers, potential updates, and adaptations to the models remain a topic of ongoing research [9]. Last year, we improved the RPCRC with a better understanding of disease burden by incorporating the latest International Society of Urological Pathology (ISUP) grading system, and invasive cribriform tumor (CR) and intraductal carcinoma (IDC) growth patterns. We found a considerable decrease in unnecessary prostate biopsies and overdiagnosis of potentially indolent disease [10]. These growth patterns subdivide indolent and csCaP within the Gleason score 7 (3 + 4) CaP [11]. Also, the 2014 ISUP grading of CaP provided more accurate stratification of tumors than the previous system [12]. The ISUP grading Gleason score 6 and below CaP were condensed in a single ISUP grade 1 category and Gleason score 7 CaP were split into 2 categories, ISUP grade 2 for Gleason score 7 (3 + 4) and ISUP grade 3 for Gleason score 7 (4 + 3). Furthermore, now all CR and glomeruloid pattern carcinoma are included in the Gleason grade 4 pattern, whereas IDC has been excluded from grading.

In the field of novel biomarkers for CaP detection, the 4-kallikrein (4K) panel comes forth in outperforming total PSA and consists of total PSA, fPSA, intact PSA, and kallikrein-related peptidase 2 (hK2) [13]. The 4K panel, together with clinical parameters (age and outcome of digital rectal examination [DRE]) have been combined into a multivariable model (4Kscore). In several large prospective studies, the 4Kscore considerably reduced unnecessary biopsies without missing many csCaP [13-16]. These findings and continuous improvements may further improve the promoted risk-adapted strategy in the prediction of csCaP for an individual

patient [17]. The aim of this study was to evaluate the clinical impact of the 4Kscore, RPCRC, and the combination of both in predicting the improved definition of csCaP based on the latest ISUP grading system and inclusion of cribriform growth pattern in Gleason 4 CaP.

MATERIALS AND METHODS

Patients

We studied men who were biopsied due to a PSA ≥ 3.0 ng/ml in the first round of the screening program of the European Randomized Study of Screening for Prostate Cancer (ERSPC) section Rotterdam [11]. The inclusion period for the randomized screening trial was November 1993 to March 2000; the detailed protocol and ethical approval has been described previously [4]. We included 3,028 men who received sextant biopsies solely on the basis of an elevated PSA (PSA ≥3.0 ng/ml). We did not include men with PSA <3.0 ng/ml who had abnormal DRE or hypoechogenic lesion on transrectal ultrasound (TRUS). This cohort was also used in a previously published 4Kscore validation study [15]. Clinical measurements (age, PSA, DRE, and prostate volume) were obtained prospectively.

In 2015, the prostate biopsy slides were re-evaluated in order to adapt to the 2014 ISUP grading system and to record CR and/or IDC components [10, 11]. The urogenital pathologist and 2 pathologists in training were blinded to patient information and biopsy outcome. The primary outcome was detection of csCaP on biopsy defined as ISUP grade 2 CaP with CR/IDC plus all CaP with ISUP grade ≥3. Secondary outcome was the detection of indolent CaP defined as ISUP grade 1 or 2 without CR/IDC [10]. We excluded 114 men whose 4K panel could not be measured due to insufficient frozen blood samples, and 42 men whose histopathology slides could not be re-evaluated. Total of 2,872 men could be used in the analysis.

The 4K panel

The 4K panel (total PSA, free PSA, intact PSA, and human kallikrein-related 2) was measured in frozen serum samples at the Wallenberg Research Laboratories, Department of Laboratory Medicine, Lund University, University Hospital UMAS in Malmö, Sweden, using the dual-label DELFIA Prostatus total/free PSA-Assay (Perkin–Elmer, Turku, Finland) [13].

Risk prediction models

The RPCRC was applied to provide the probabilities of csCaP and indolent CaP on a biopsy. These probabilities were calculated by the developed multinomial logistic regression analysis with PSA, prostate volume estimated through DRE, DRE abnormalities, and age as predictors [10]. Since no data on prostate volume assessed with DRE was available in this

validation cohort, the available TRUS-assessed prostate volume was recoded into 3 volume classes as can be estimated by DRE [7]. TRUS-assessed volumes <30 cc were recoded as 25 cc, volumes between 30 and 50 cc as 40 cc, and volumes >50 cc as 60 cc. The effect of interobserver variability of DRE volume estimation on the performance of the RPCRC was externally validated and was negligible due to substantial agreement in DRE volume estimation. In addition, there was a good correlation between the TRUS-assessed volume and DRE-estimated volume [18].

The 4Kscore is an algorithm constructed from the 4K panel and clinical parameters. Originally, a laboratory model and clinical model were developed by Vickers et al. [13]. The clinical model consists of the 4K panel, age, and DRE outcome (4Kscore). A blinded dataset for outcome has been sent to the 4Kscore developers in order to receive the 4Kscore probability of csCaP on biopsy. To update the RPCRC with the 4Kscore, we first recalibrated the 4Kscore by re-estimating the intercept and slope of the linear predictor, and, subsequently, we added the 4Kscore as a predictor [19].

Statistical analysis

The predictive performance of each risk model was evaluated according to the area under the receiver operator characteristic curve (AUC). Correction for overestimation was done by bootstrapping techniques using 1,000 samples. Differences in AUCs were tested after calculation of the standard error of the AUC with the DeLong method [20]. Calibration was assessed by grouping men by deciles of absolute risk. The observed and expected counts of incident CaP cases in each decile were compared for deviance and significance calculated according to the Hosmer–Lemeshow statistic [21]. Finally, reduction of unnecessary biopsy by 4Kscore, RPCRC, and combination of both was assessed with decision curve analysis and the net benefit (NB) formula [22]. The additional reduction of unnecessary biopsy without missing any csCaP for the 4Kscore, RPCRC, and the 4Kscore-RPCRC combination compared to the defaults strategy (biopsy PSA ≥ 3.0 ng/ml) was calculated with the net avoided false positives formula per 100 men at risk: ΔNB/odds(T)*100, where ΔNB is the NB difference between the models and the default strategy; T = risk threshold. By inversing ΔNB/odds(T), it was possible to assess whether the reduction of unnecessary biopsies would outweigh the costs associated using the models [23]. Baseline characteristics were tested with the nonparametric Kruskal–Wallis test for continuous variables, and with chi-square for categorized values. *P* values < 0.05 were taken to indicate statistical significance. Statistical computations were performed with R, version 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria).

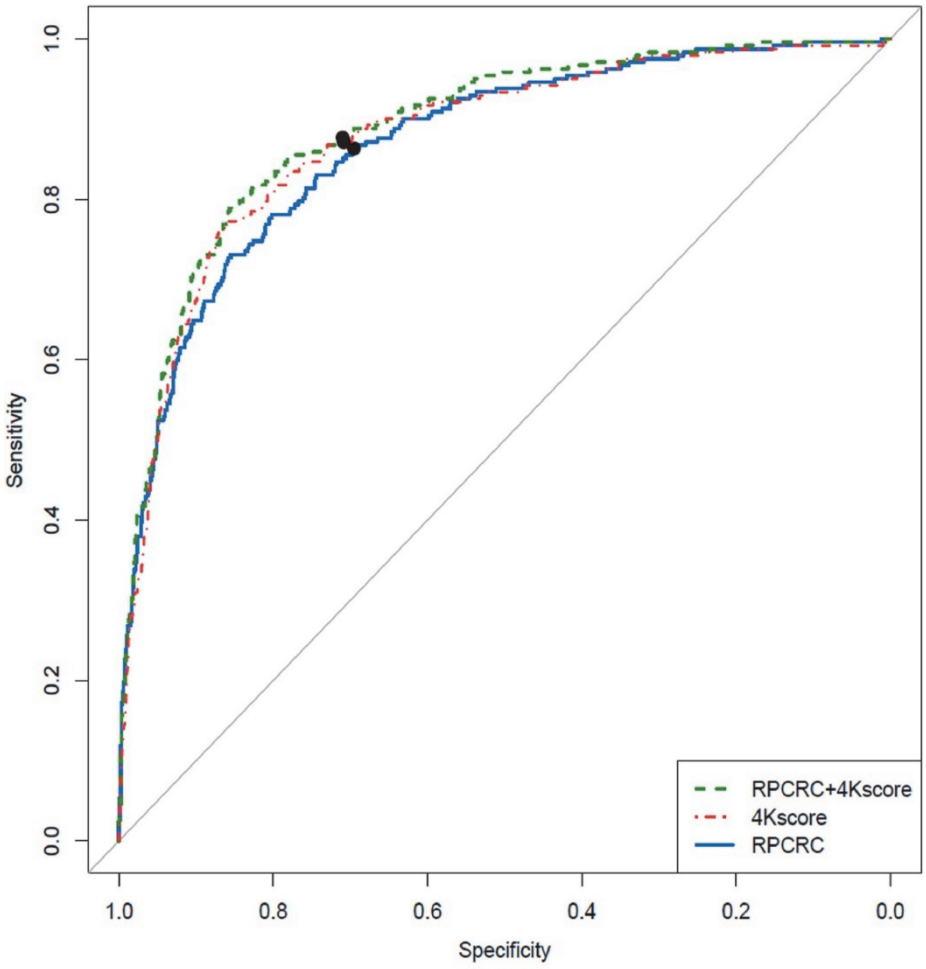
RESULTS

Baseline characteristics of the 2,872 men stratified by biopsy outcome are presented in Table 1. A total of 578 (20%) indolent CaP cases and 242 (8%) csCaP cases were detected. Men with csCaP had higher 4K values ($P < 0.01$), smaller prostate volume ($P < 0.01$), and more abnormal digital rectal findings ($P < 0.01$) compared to men without CaP. The 4Kscore and RPCRC predicted csCaP with a similar high AUC (0.88 vs. 0.87; $P = 0.41$). The 4Kscore-RPCRC combination improved AUC to 0.89 compared to 4Kscore ($P < 0.01$) and RPCRC ($P < 0.01$), as shown in Fig. 1. At a 5% risk threshold, the 4Kscore-RPCRC combination had a sensitivity of 88% and specificity of 71%. Calibration was assessed with the Hosmer–Lemeshow test and indicated a good calibration of the combined 4Kscore with RPCRC ($P = 0.09$), see Appendix Fig. A.1.

Table 1. Patient and tumor characteristics stratified to groups without and with indolent or clinically significant prostate cancer

Characteristic	No CaP n = 2,052 (72%)	Indolent CaP n = 578 (20%)	Clinically significant CaP n = 242 (8%)
Age, y, median (IQR)	66 (62–70)	66 (62–70)	69 (64–72)
Kallikrein panel, ng/ml, median (IQR)			
PSA	4.8 (3.9–6.4)	5.5 (4.0–7.9)	9.8 (4.9–17.0)
Free PSA	1.1 (0.8–1.5)	1.0 (0.7–1.4)	1.1 (0.7–1.9)
Intact PSA	0.5 (0.4–0.7)	0.5 (0.4–0.7)	0.7 (0.4–1.1)
hK2	0.07 (0.05–0.09)	0.08 (0.05–0.11)	0.11 (0.07–0.18)
Prostate volume, ml, median (IQR)	48 (38–63)	39 (29–48)	36 (29–48)
Abnormal DRE, *n* (%)	424 (21)	201 (35)	167 (69)
ISUP, *n* (%)			
Grade 1	–	364 (63)	–
Grade 2	–	214 (37)	48 (20)
Grade 3	–		82 (34)
Grade 4 and 5	–		112 (46)
CR/IDC presence, *n* (%)	–	–	158 (65)

IQR = Interquartile range; Indolent CaP = International Society of Urological Pathology (ISUP) grade 1 or ISUP grade 2 without presence of CR/IDC; clinically significant CaP = ISUP grade ≥3 or ISUP grade 2 with presence of CR/IDC; DRE = digital rectal examination



	Threshold 5%			AUC model comparison	
Models	Sensitivity	Specificity	AUC	RPCRC p-value	4Kscore p-value
RPCRC	0.86	0.70	0.868	-	0.43
4Kscore	0.87	0.71	0.876	0.43	-
RPCRC + 4Kscore	0.88	0.71	0.888	<0.01	<0.01

Figure 1. Receiver operating characteristic for RPCRC, 4Kscore, and the combined risk calculators predicting csCaP; the black dots display the 5% risk thresholds; sensitivity, and specificity are displayed. AUC = area under the curve of the receiver operating characteristic; 4K = 4-kallikrein; RPCRC = Rotterdam Prostate Cancer Risk Calculator.

Table 2. Performed and reduced numbers of biopsy, delayed prostate cancer diagnosis, and net true reduction in biopsies for the RPCRC, 4Kscore, and combination of both models compared to the standard strategy (PSA ≥ 3.0 ng/ml), with different thresholds for csCaP

Threshold	Number of biopsies		Number of indolent cancers		Number of high-grade cancers		Additional net true reduction of biopsies compared to	
Clinically significant CaP	Performed	Reduced	Detected	Not detected (% reduced biopsies)	Detected	Missed diagnosis (% reduced biopsies)	PSA ≥ 3.0 (×100 men)	RPCRC (×100 men)
Biopsy all	2,872	0	578	0	242	0	-	-
4%								
RPCRC-Cribriform	1,142	1,730 (68%)	317	261 (15%)	213	29 (2%)	35.0	Ref.
4Kscore	1,196	1,676 (58%)	370	208 (12%)	218	24 (1%)	37.5	2.5
4K+RPCRC-Cribriform	1,151	1,721 (60%)	361	217 (13%)	219	23 (1%)	39.9	4.9
5%								
RPCRC-Cribriform	1,011	1,861 (65%)	285	293 (16%)	209	33 (2%)	41.9	Ref.
4Kscore	978	1,894 (66%)	322	256 (14%)	211	31 (2%)	44.4	2.5
4K+RPCRC-Cribriform	975	1,897 (66%)	318	260 (14%)	212	30 (2%)	45.2	3.3
7.5%								
RPCRC-Cribriform	772	2,100 (73%)	226	352 (17%)	189	53 (3%)	48.5	Ref.
4Kscore	724	2,148 (75%)	246	332 (15%)	198	44 (2%)	53.4	4.9
4K+RPCRC-Cribriform	711	2,161 (75%)	245	333 (15%)	200	42 (2%)	55.7	7.2
10%								
RPCRC-Cribriform	613	2,259 (79%)	182	396 (18%)	178	64 (3%)	56.4	ref
4Kscore	582	2,290 (80%)	203	375 (16%)	187	55 (2%)	60.6	4.2
4K+RPCRC-Cribriform	576	2,296 (80%)	190	388 (17%)	191	51 (2%)	62.2	5.8

Bold values indicate the risk thresholds for detection of csCaP with the different models. 4K = 4-kallikrein; PSA = Prostate Specific Antigen; RPCRC = Rotterdam Prostate Cancer Risk Calculation.

DISCUSSION

As compared to a "biopsy all men with a PSA ≥ 3.0 ng/ml" strategy, our results show that basing the decision to biopsy on the RPCRC or the 4Kscore gives an equal reduction (approximately two-third) in the number of biopsies and the rate of overdiagnosis. The combination of 4Kscore and RPCRC slightly further reduced the number of unnecessary biopsies without missing additional csCaP diagnoses.

As both the 4Kscore and the RPCRC showed similar reduction of unnecessary biopsies and overdiagnosis of indolent CaP in our study cohort, other aspects should be considered when evaluating clinical usefulness. Factors that are important to consider are e.g., burden to the patient, availability of the test, and costs. Both risk calculators use PSA, age, and DRE findings. The risk calculators differ in estimation of prostate volume: the 4Kscore uses the 4K panel (total PSA, free PSA, intact PSA, and human kallikrein-related 2) as a proxy for prostate volume [15], where the RPCRC estimates the prostate volume based on DRE (prostate volume is categorized as <30, 30–49 and ≥50 cm3) [8]. Here, it should be noted that a DRE volume estimation is a subjective procedure while PSA (sub forms) measurements are objective and in principle independent of the experience of the executor. Since both risk calculators require blood and DRE, the burden to the patient is therefore similar for both approaches and outpatient clinic costs are comparable to both risk calculator approaches. In terms of test-availability the 4K panel is not available in certain areas (e.g., Europe) and requires sending the blood to an external laboratory for analysis, which involves logistics and additional costs. The price of the 4K panel is estimated to range from $400 to $1100. The RPCRC is freely available on the internet or purchasable as an app for 1.99 dollar.

Besides comparison of the models we fused both models to optimize csCaP prediction, and found that the 4Kscore needs to be applied to 30 men to avoid 1 man getting an unnecessary biopsy compared to the RPCRC alone. A detailed cost effectiveness study was outside the scope of this research, however, on estimation prediagnostic work-up would increase up to $12,000 to $33,000 to avoid 1 additional unnecessary biopsy. A recent cost-effectiveness study in the United States showed a reduction in health care costs by applying 4Kscore before biopsy compared to biopsy all patients referred to a urologist with suspicion of having CaP in a theoretical cohort [24]. Most likely, a larger cost reduction could be achieved when applying a freely available risk stratification tool. In addition, it would be of value to perform cost-effectiveness studies within a primary care setting where initial risk stratification with the RPCRC could be (and actually is in the Netherlands [25]) applied.

Strengths of our study are the large population-based sample size, prospectively collected measurements of risk factors allowing for updating, and a contemporary pathology review

including the latest ISUP scoring system. The ERSPC section Rotterdam cohort on which both models were originally developed was used to enable a fair comparison between the two risk prediction models [13]. To our knowledge, there is no other large population-based sample size available including the detailed pathological grading and data on the 4K pattern. Limitations of our study are that we did not have the actual DRE estimated prostate volume available in this cohort, the DRE estimated prostate volume was derived using the TRUS based volume. The DRE based volume estimate approach is externally validated in a clinical setting and showed good concordance [18]. In addition, using volume classes has no effect in discriminatory ability as was confirmed in the development of models with a urinary molecular biomarker-based score [26]. The ERSPC section Rotterdam cohort represents two decades ago first-time screened men, who are predominantly white and received sextant biopsies. Sextant biopsies are known to detect less CaP than the present standard of 12-core based transrectal ultrasound biopsy procedures [27], thus it might have under-represented the CaP detection rate. However, in multiple contemporary clinical settings with 12-core TRUS biopsies, the RPCRC showed good predictive performance [28], [29]. A comparable situation holds for the 4Kscore which is also partly developed on similar cohorts reflecting old practices, including, next to the currently used Dutch ERSPC cohort, the Swedish ERSPC data [13]. Also, the 4Kscore performs well after external validation in a contemporary clinical setting [30]. Although our comparison is based on prediction tools that are (partly) based on cohorts from two decades ago, this comparison is still relevant for today's clinical practice. This is even confirmed by external validation in cohorts where multi-parametric Magnetic resonance imaging (mpMRI) targeted biopsies were applied [31]. Individualized risk assessment on having a biopsy detectable CaP including mpMRI Prostate Imaging Reporting and Data System (PI-RADS) score and clinical data can result in a considerable reduction of unnecessary biopsies [32]. Hence, biomarkers and mpMRI results could be combined to optimize upfront risk prediction. It must however be noted that long-term outcomes like e.g., metastatic disease and/or CaP death are not available for a detection pathway driven on individual risk and mpMRI targeted biopsy. The European Association of Urology guidelines recommend that a clinical risk prediction tool should be incorporated in the decision-making process, as the “one-size-fits-all” approach with a PSA cutoff does not provide a good balance between reduction of CaP morbidity, mortality, and overdiagnosis of indolent CaP [17]. A recent head-to-head comparison of the most well-known prediction tools showed that the ERSPC RPCRC is superior in identifying those men at risk for csCaP compared to Prostate Cancer Prevention Trial and Sunnybrook risk calculators [33]. Continuous updating of these existing prediction models will refine the balance between harms and benefit, as previously demonstrated with further refining the definition of csCaP [10], by incorporating the latest ISUP grading system [12]. Still future research is necessary by combining novel biomarkers, MRI findings, and the latest Gleason grading modifications (cribriform architecture) to fully assess the potential of the currently available prediction tools in contemporary clinical trials

to reduce the detection of those CaP that will never become life-threatening as well as those CaP deemed suitable for active surveillance [11].

CONCLUSION

The RPCRC and 4Kscore had similar reductions of unnecessary biopsies and overdiagnosis of indolent CaP. Combination of both models slightly further reduced unnecessary biopsies. Given that the improvement in clinical impact was marginal, adding additional risk factors and biomarkers associated with CaP risk remains a tradeoff where aspects such as costs and patient burden are important considerations. More research is needed to validate the updated model in independent, contemporary, and various populations.

ACKNOWLEDGMENTS

We would like to thank A. Vickers and H. Lilja for providing the 4Kscore data.

REFERENCES

1. F.H. Schröder, J. Hugosson, M.J. Roobol, T.L.J. Tammela, M Zappa, V Nelen, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet, 384 (2014), pp. 2027-2035
2. S.M. Bruinsma, C.H. Bangma, P.R. Carroll, M.S. et al. Active surveillance for prostate cancer: a narrative review of clinical guidelines. Nat Rev Urol, 13 (2016), pp. 151-167
3. L.J. Esserman, I.M. Thompson, B. Reid, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol, 15 (2014), pp. e234-e242
4. M.J. Roobol, E.W. Steyerberg, R. Kranse, T. Wolters, R.C. van den Bergh, C.H. Bangma, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. Eur Urol, 57 (2010), pp. 79-85
5. D.P. Ankerst, J. Hoefler, S. Bock, P.J. Goodman, A. Vickers, J. Hernandez, et al. Prostate cancer prevention trial risk calculator 2.0 for the prediction of low-vs high-grade prostate cancer. Urology, 83 (2014), pp. 1362-1367
6. K.S. Louie, A. Seigneurin, P. Cathcart, P. Sasieni. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. Ann Oncol, 26 (2015), pp. 848-864
7. M.J. Roobol, H.A. van Vugt, S. Loeb, X. Zhu, M. Bul, C.H. Bangma, et al. Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators. Eur Urol, 61 (2012), pp. 577-583
8. M.J. Roobol, F.H. Schröder, J. Hugosson, J.S. Jones, M.W. Kattan, E.A. Klein, et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group. World J Urol, 30 (2012), pp. 149-155
9. D.K. Pruthi, D.P. Ankerst, M.A. Liss. Novel definitions of low-risk and high-risk prostate cancer: implications for the European Randomized Study of Screening for Prostate Cancer risk assessment tool. Eur Urol, 72 (2017), pp. 52-53
10. M.J. Roobol, J.F.M. Verbeek, T. van der Kwast, I.P. Kummerlin, C.F. Kweldam, G. van Leenders. Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer risk calculator for initial prostate biopsy by incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. Eur Urol, 72 (2017), pp. 45-51
11. C.F. Kweldam, I.P. Kummerlin, D. Nieboer, E.I. Verhoef, E.W. Steyerberg, T.H. van der Kwast, et al. Disease-specific survival of patients with invasive cribriform and intraductal prostate cancer at diagnostic biopsy. Mod Pathol, 29 (2016), pp. 630-636
12. J.I. Epstein, L. Egevad, M.B. Amin, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. Am J Surg Pathol, 40 (2016), pp. 244-252
13. A.J. Vickers, A.M. Cronin, G. Aus, C.G. Pihl, C. Becker, K. Pettersson, et al. A panel of kallikrein markers can reduce unnecessary biopsy for prostate cancer: data from the European Randomized Study of Prostate Cancer Screening in Goteborg. BMC Med, 6 (2008), p. 19
14. S. Carlsson, A. Maschino, F. Schröder, C. Bangma, E.W. Steyerberg, T. van der Kwast, et al. Predictive value of four kallikrein markers for pathologically insignificant compared with aggressive prostate cancer in radical prostatectomy specimens: results from the European Randomized Study of Screening for Prostate Cancer section Rotterdam. Eur Urol, 64 (2013), pp. 693-699

15. A.J. Vickers, A.M. Cronin, M.J. Roobol, C.J. Savage, M. Peltola, K. Pettersson, et al. A four-kallikrein panel predicts prostate cancer in men with recent screening: data from the European Randomized Study of Screening for Prostate Cancer, Rotterdam. Clin Cancer Res, 16 (2010), pp. 3232-3239

16. A. Benchikh, C. Savage, A. Cronin, G. Salama, A. Villers, H. Lilja, et al. 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, 10 (2010), p. 635

17. N. Mottet, J. Bellmunt, M. Bolla, E. Briers, M.G. Cumberbatch, M. De Santis, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol, 71 (2017), pp. 618-629

18. N. Pereira-Azevedo, I. Braga, J.F. Verbeek, L. Osorio, V. Cavadas, A. Fraga, et al. Prospective evaluation on the effect of interobserver variability of digital rectal examination on the performance of the Rotterdam prostate cancer risk calculator. Int J Urol, 24 (2017), pp. 826-832

19. E.W. Steyerberg, G.J. Borsboom, H.C. van Houwelingen, M.J. Eijkemans, J.D. Habbema. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med, 23 (2004), pp. 2567-2586

20. E.W. Steyerberg, A.J. Vickers, N.R. Cook, T. Gerds, M. Gonen, N. Obuchowski, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology, 21 (2010), pp. 128-138

21. D.W. Hosmer, T. Hosmer, S. Le Cessie, S. Lemeshow. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med, 16 (1997), pp. 965-980

22. A.J. Vickers, E.B. Elkin. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making, 26 (2006), pp. 565-574

23. B. Van Calster, L. Wynants, J.F.M. Verbeek, J.Y. Verbakel, E. Christodoulou, A.J. Vickers, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol, 74 (2018), pp. 796-804

24. J.D. Voigt, Y. Dong, V. Linder, S. Zappala. Use of the 4Kscore test to predict the risk of aggressive prostate cancer prior to prostate biopsy: overall cost savings and improved quality of care to the us healthcare system. Rev Urol, 19 (2017), pp. 1-10

25. D.F. Osses, A.R. Alberts, G.C.F. Bausch, M.J. Roobol. Multivariable risk-based patient selection for prostate biopsy in a primary health care setting: referral rate and biopsy results from a urology outpatient clinic. Transl Androl Urol, 7 (2018), pp. 27-33

26. L. Van Neste, R.J. Hendriks, S. Dijkstra, G. et al. Detection of high-grade prostate cancer using a urinary molecular biomarker-based risk score. Eur Urol, 70 (2016), pp. 740-748

27. V. Scattoni, C. Maccagnano, U. Capitanio, A. Gallina, A. Briganti, F. Montorsi. Random biopsy: when, how many and where to take the cores? World J Urol, 32 (2014), pp. 859-869

28. H.A. van Vugt, R. Kranse, E.W. Steyerberg, H.G. van der Poel, M. Busstra, P. Kil, et al. Prospective validation of a risk calculator which calculates the probability of a positive prostate biopsy in a contemporary clinical cohort. Eur J Cancer, 48 (2012), pp. 1809-1815

29. M. Gayet, C.K. Mannaerts, D. Nieboer, H.P. Beerlage, H. Wijkstra, P.F.A. Mulders, et al. Prediction of prostate cancer: external validation of the ERSPC risk calculator in a contemporary Dutch clinical cohort. Eur Urol Focus, 4 (2018), pp. 228-234

30. D.J. Parekh, S. Punnen, D.D. Sjoberg, S.W. Asroff, J.L. Bailen, J.S. Cochran, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol, 68 (2015), pp. 464-470

31. Alberts AR, Roobol MJ, Verbeek JFM, Schoots IG, Chiu PK, Osses DF, et al. Prediction of high-grade prostate cancer following multiparametric magnetic resonance imaging: improving the Rotterdam European randomized study of screening for prostate cancer risk calculators. Eur Urol. 2019 Feb;75(2):310-318.

32. P.J. van Leeuwen, A. Hayen, J.E. Thompson, D. Moses, R. Shnier, M. Bohm, et al. A multiparametric magnetic resonance imaging-based risk model to determine the risk of significant prostate cancer prior to biopsy. BJU Int, 120 (2017), pp. 774-781

33. N. Pereira-Azevedo, J.F.M. Verbeek, D. Nieboer, C.H. Bangma, M.J. Roobol. Head-to-head comparison of prostate cancer risk calculators predicting biopsy outcome. Transl Androl Urol, 7 (2018), pp. 18-26

SUPPLEMENTARY MATERIALS

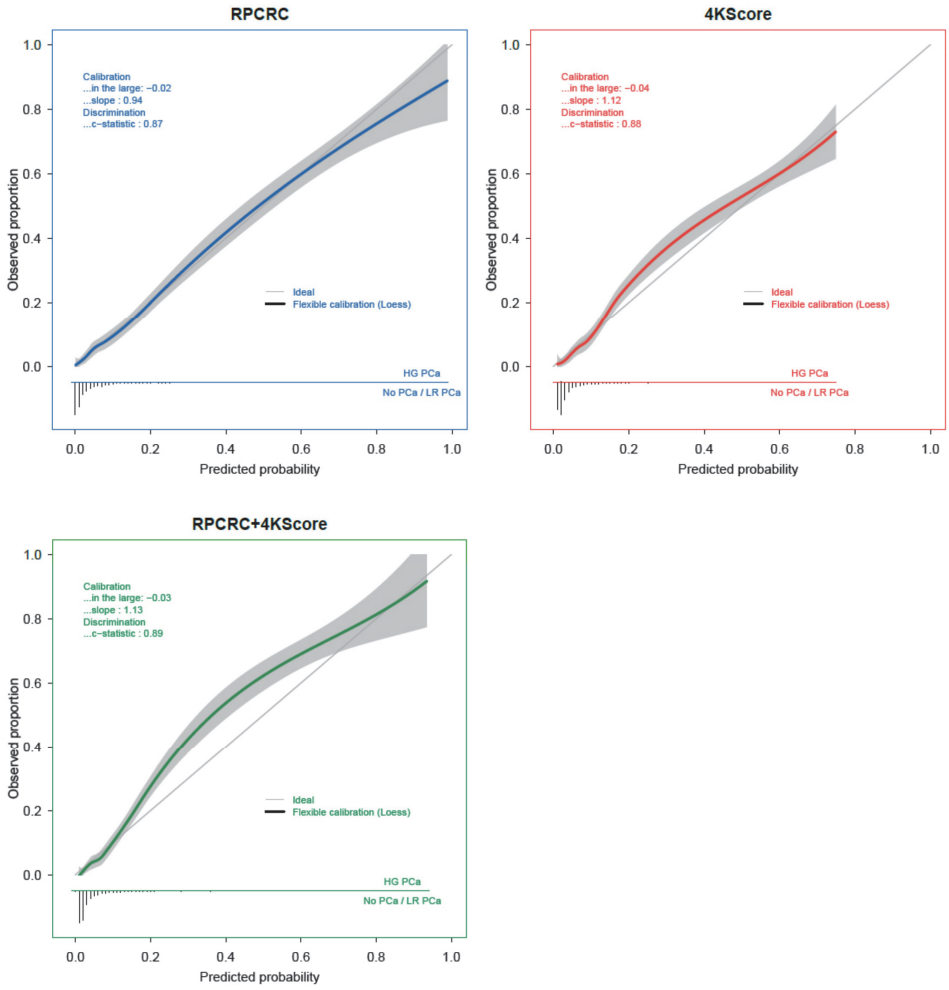


Figure A.1. Calibration plots predicting csCaP for the 3 risk calculators.

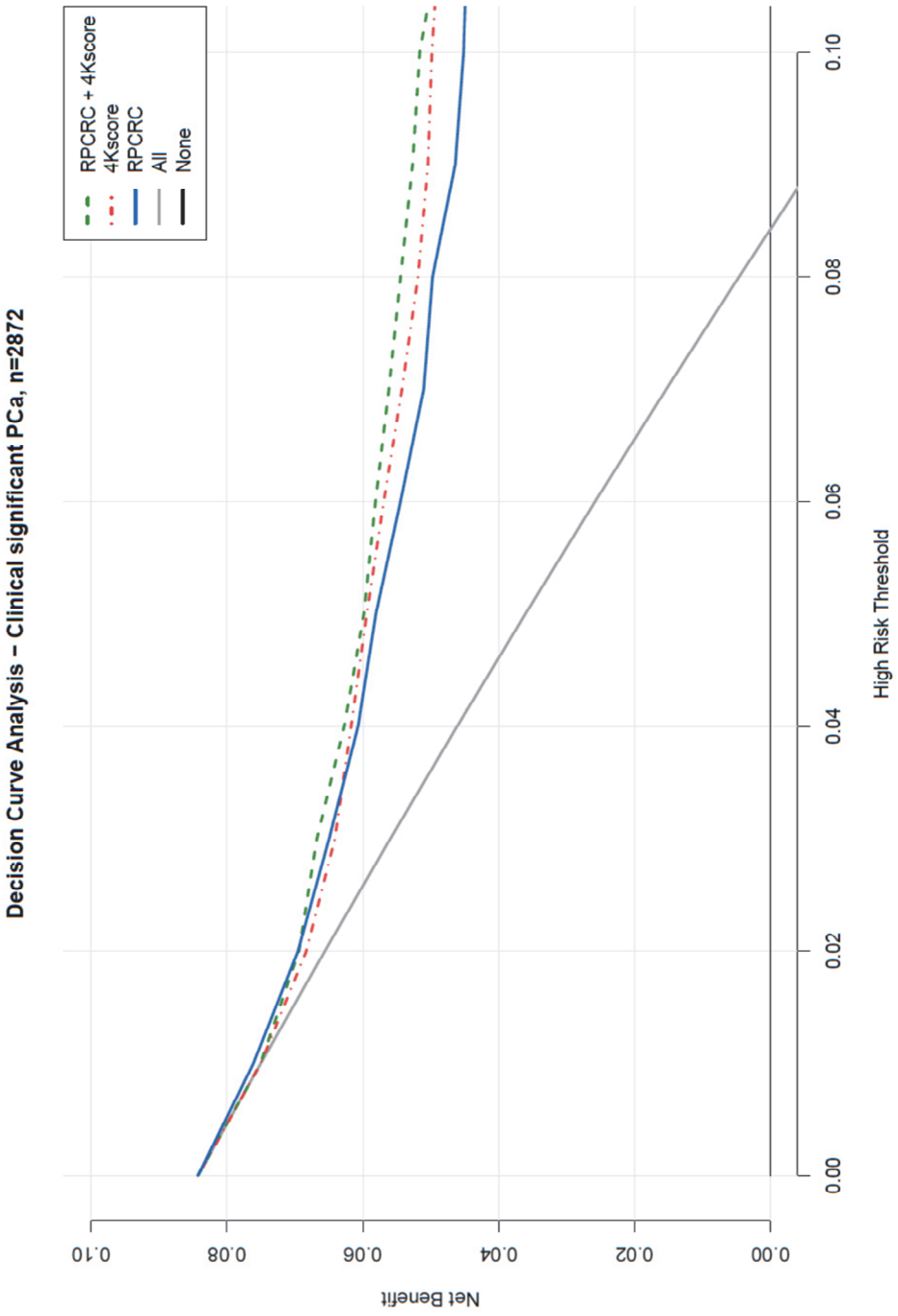
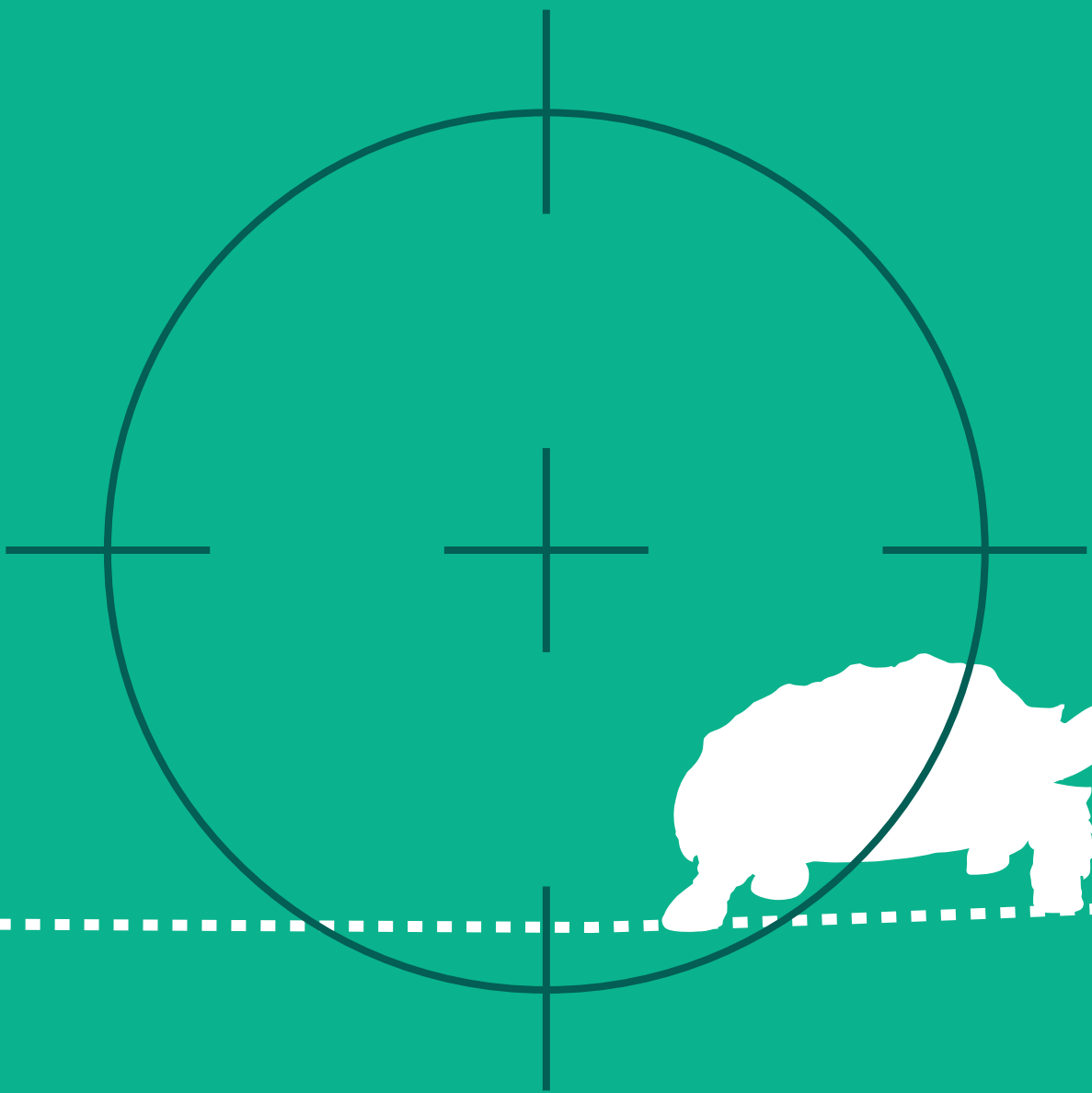


Fig. A.2. Decision curves showing the NB value when applying the RPCRC, 4Kscore, and combined risk calculator in men at risk for csCaP.



CHAPTER 7

A tool for shared decision making on referral for prostate biopsy in the primary care setting: Integrating risks of cancer with life expectancy

Verbeek JFM
Nieboer D
Parker C
Kattan MW
Steyerberg EW
Roobol MJ

Published in Special Issue Risk-stratified Cancer Screening
J Pers Med. 2019 Apr 22;9(2). pii: E19

ABSTRACT

Prostate cancer (PCa) testing involves a complex individually based decision-making process. It should consider competing risks from other comorbidities when estimating a survival benefit from the early detection of clinically significant (cs)PCa. We aimed to develop a prediction tool that provides concrete advice for the general practitioner (GP) on whether to refer a man for further assessment. We hereto combined the probability of detecting csPCa and the potential overall survival benefit from early detection and treatment. The PCa detection probabilities were derived from 3616 men enrolled in the Dutch arm of the European Randomized Study of Screening for Prostate Cancer (ERSPC). Survival estimates were derived from 19,834 men from the Surveillance, Epidemiology, and End Results (SEER) registry, ERSPC, and Dutch life tables. Treatment benefit was estimated from the Prostate Cancer Intervention versus Observation Trial (PIVOT, n = 731). The prediction of csPCa detection was based on prostate-specific antigen (PSA), age, %freePSA, and digital rectal examination (DRE). The life expectancy (LE) for patients with PCa receiving no treatment was adjusted for age and Charlson comorbidity index. A negative impact on LE and treatment benefit was found with higher age and more comorbidity. The proposed integrated approach may support triage at GP practices, as PCa is a heterogeneous disease in predominantly elderly men.

INTRODUCTION

Prostate-specific antigen- (PSA) based screening for prostate cancer (PCa) can reduce PCa mortality, as has been demonstrated in a large-scale European randomized screening trial [1]. However, PSA-based screening also results in the detection of considerable numbers of indolent PCa due to lack of risk stratification and the random method of sampling. This results in over-diagnosis and overtreatment of clinically harmless PCa negatively affecting the harm–benefit ratio [2]. Therefore, referral for further testing should only be applied to patients with high risk of metastasis and cancer-related mortality. However, this ideal risk stratification is not yet feasible, even with the use of novel techniques such as imaging and contemporary biomarkers. The U.S. Preventative Services Task Force, European Association of Urology (EAU), and American Urological Association (AUA) guidelines recommend that men aged 55–69 years should be informed about the benefits and the harms of screening, and PSA testing should be offered only after informed choice [3–5]. For most men, PCa screening starts with a visit to the general practitioner (GP). It is the GP's task/challenge to guide men and to identify men who can benefit from early detection and treatment. To assist (future) patients and physicians in interpreting the clinical significance of PSA levels, multivariable PCa risk calculators (RC) have been developed that estimate the probability of detecting potentially aggressive) PCa if referred for prostate biopsy. These RCs improve predictions by including other relevant information, such as age or family history, in addition to PSA levels [6–8]. However, these PCa RCs do not include a patient's characteristics, e.g., life expectancy (LE) and long-term effects of treatment. These are relevant factors, since risk of experiencing harm from a potentially aggressive PCa is likely to be offset, to some extent, by a reduced LE for older men [9]. To obtain insight in these competing risks, they need to be quantified and modeled. The aim of this study is to provide a tool suitable for use in primary care that, on the basis of readily available information, can assess the risk of having a potentially aggressive PCa in the context of a man's LE, and in addition, quantifies potential treatment benefit.

MATERIALS AND METHODS

Several aspects should be taken into account in a shared decision-making process to refer for a biopsy: first, an individual's current risk of having clinically significant PCa (csPCa) [International Society of Urological Pathology (ISUP) grade ≥2]; second, his LE in the absence of csPCa; third, his LE in the case of undetected and untreated csPCa; and lastly, how much benefit could be gained from treatment in the case of csPCa diagnosis? Since there is no single dataset available comprehensive enough to simultaneously assess these individual probabilities, multiple data sources were used for the development of the proposed tool. Figure 1 provides an overview of the different prediction models and their underlying

sources. To summarize, the model predicting the presence of csPCa at the time of biopsy was based on the Dutch arm of the European Randomized Study of Screening for Prostate Cancer (ERSPC). Estimates on LE of men diagnosed with csPCa receiving no active treatment were based on the Surveillance, Epidemiology, and End Results (SEER) registry. The benefit of active treatment was estimated from the Prostate Cancer Intervention versus Observation Trial (PIVOT). To predict the LE for men with csPCa receiving active treatment, the treatment benefit from the PIVOT was added to the LE prediction from the SEER. Estimates on LE for men without csPCa were predicted using data from the Dutch ERSPC. The outcomes of the different prediction models are displayed in an easy-to-read format that enables evaluation of csPCa risk in the context of a man's LE and treatment effect. Final recommendations for GPs formulated as "no referral needed" or "refer to a urologist" were based on consensus of risk thresholds by the Prostate Cancer United Kingdom Prostate Risk Working Group (PCUK-RWG) [10], taking into account the probabilities for having csPCa on a current biopsy (>5–10%), life expectancy (>10–15 years), and treatment benefit (1–2 years additional gain in overall survival). If the calculated risk is below the lower limit, the advice is not to refer. If the risk is within range (see Figure 1), the patient's preferences can be dictated. If calculated risks are above the upper limit, the patient should definitely be referred. It should be noted that referral to secondary care is also indicated when multiple criteria are above the given range. Here, the risk of csPCa and the potential treatment benefit should be leading, even with an LE estimated to be below 10 years. For example, if a patient would have an elevated risk of having csPCa when biopsied, an estimated LE of nine years without csPCA, but a potential treatment benefit of three years, he should be referred for biopsy, and when the suspicion of csPCa is confirmed, he should be actively treated. The analysis for each prediction model is described in detail below.

The risk of having a biopsy-detectable csPCa is based on 3616 men who received transrectal ultrasound-guided sextant biopsies in the first screening round of the Dutch arm of the ERSPC [7]. Only variables to which a GP has easy access were included in the analyses, i.e., age at time of biopsy, PSA (two log centered), %freePSA (freePSA/total PSA; two log centered), results of the digital rectal examination (DRE) including a rough estimate of prostate volume (PV) estimated during DRE (25, 40, or 60 cc; two log centered [11]), family history, and the International Prostate Symptom Score (IPSS). These predictors were combined in a series of logistic regression models in which the discriminative ability of each model was assessed. First, the model was fitted to all observations in the given set, and the concordance index was calculated. Second, a dataset was formed by bootstrapping with 1000 samples in which the model was again developed and then validated based on the original data. The difference in performance between the original and the bootstrapped data was the estimated "optimism". The models with the highest concordance index after correction for optimism were selected.

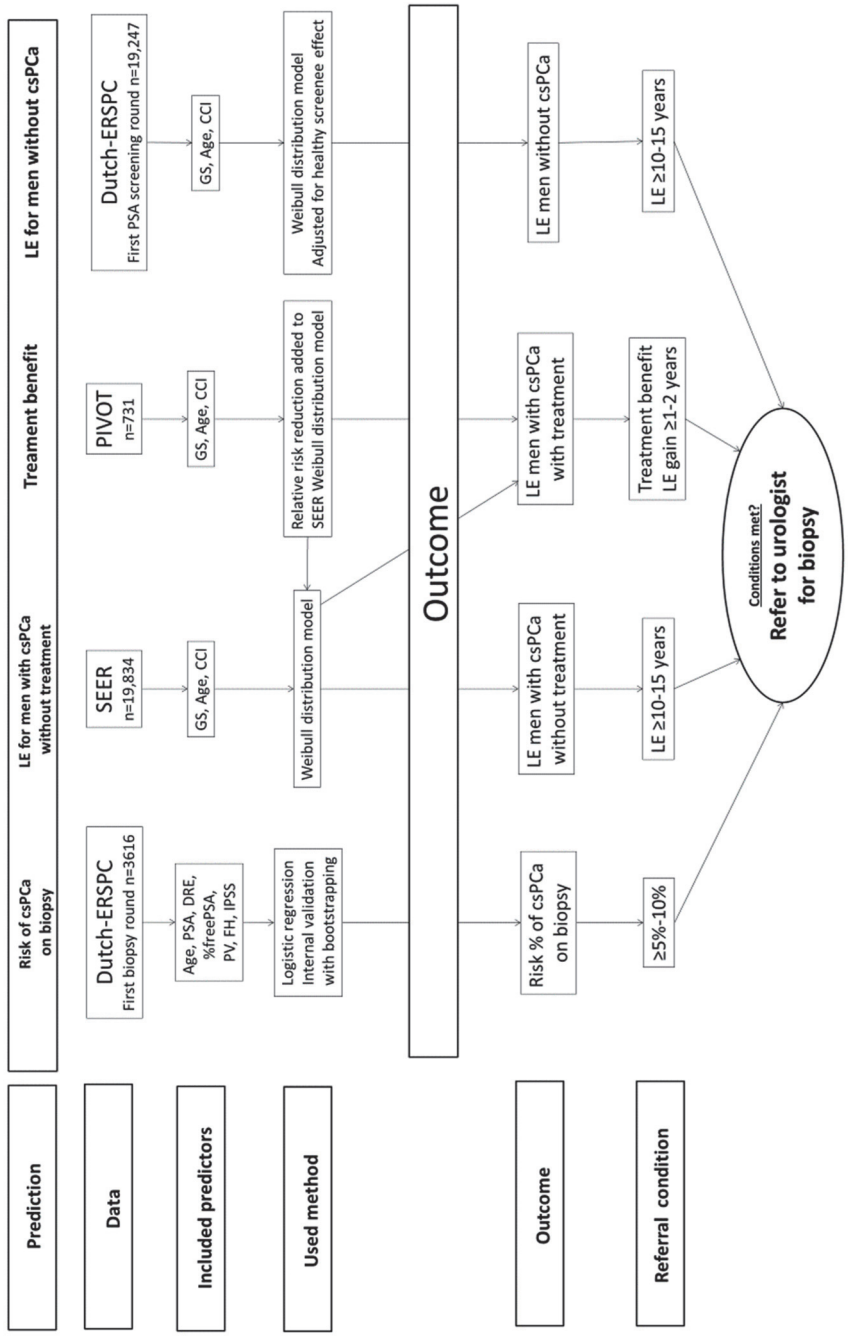


Figure 1. Flowchart of development of prediction model predicting risk of csPCa on current biopsy, overall life expectancy, and treatment benefit for each individual patient. LE: life expectancy, csPCa: clinically significant prostate cancer.

As the concordance index does not reflect calibration, the Index of Prediction Accuracy (IPA) considering both discrimination and calibration was calculated; a higher IPA indicated more accurate predictions [12]. The clinical utility of the models was expressed with net benefit (NB) by summing the benefits (true positive biopsies) and subtracting the harms (unnecessary biopsy). The harms were weighted by a factor related to the relative harm of a missed cancer versus unnecessary biopsies [13]. This weighting was derived from the threshold probability for csPCa at which a patient would opt for a biopsy (range considered 3–10%) and were displayed in a decision curve analysis graph. A model was considered to be clinically useful if its NB was higher than the default strategy (biopsy if PSA ≥ 3.0 ng/mL).

The LE for men with csPCa without receiving treatment was estimated based on the SEER program [14]. The SEER consisted of 19,639 men (age ≥65) diagnosed in the period from 1 January 1992 to 31 December 2005 and 195 men (age <65) diagnosed between 1 January 1971 and 31 December 1984 [15,16]. The SEER reported overall mortality outcomes for Gleason Score (Gleason Score 5–7 and Gleason 8–10), age (55–59, 60–64, 65–69, 70–74, 75+) and Charlson comorbidity index (0, 1, ≥2). These survival curves were approximated using a Weibull distribution. To estimate LE for men with Gleason 3 + 4 or higher (ISUP grade ≥2), the Gleason Score distribution from the Dutch ERSPC was used to adjust the SEER's reported Gleason Score distribution [17].

Also, a relative effect of 0.79 on PCa mortality was applied to the pre-PSA era SEER cohort to include the reduction in PCa mortality by the introduction of PSA [16,18]. The reported SEER's outcomes and adjustments were fitted in a model with a Weibull distribution to predict individual LE for men with csPCa receiving no treatment.

The treatment benefit of csPCa was based on the PIVOT [19]. The PIVOT is a randomized trial comparing treatment effect of radical prostatectomy versus watchful waiting in 731 men with localized prostate cancer diagnosed in the era of PSA testing. The relative effect for all-cause mortality for treatment versus no active treatment was extracted, and the life expectancy for men with csPCa and treatment were estimated using the survival curves, which were adjusted using the relative treatment effect from the PIVOT trial. Besides the PIVOT study, other randomized clinical trials comparing PCa treatment with observation include the Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4) and the Prostate Testing for Cancer and Treatment (ProtecT) trial [20,21]. These trials have similar long follow-up. A sensitivity analysis was performed to compare the LEs after treatment for men with csPCa using the relative treatment effects from the PIVOT, SPCG-4, and ProtecT.

The LE for men without csPCa was estimated using data of the Dutch ERSPC [1,22]. In the period 1993–1999, a total of 21,210 men (age 54–74) were randomized to the screening arm; 19,970 men had a PSA test at the first screening round. We excluded men diagnosed with csPCa in the first round (n = 313) and men with life-threatening malignancies (n = 410), such as lung cancer, colon cancer, and leukemia. These patients should not be tested for prostate cancer, since the likelihood that they would benefit from an early PCa diagnosis is low [23].

Skin cancer was not an exclusion criterion. This led to a total of 19,247 men available for the prediction of LE for men without csPCa. Survival and follow-up time in months since time of first visit, survival status (dead or alive), age at visit, and Charlson comorbidity index were entered in a Weibull distribution model to calculate the LE for an individual without csPCa. Data for survival status was obtained by linkage with national registries (Central Bureau for Statistics, 2015). Charlson comorbidity index was missing in 158 cases and was imputed using multiple imputation with the chained equations procedure and predictive mean matching [24]. As men with a healthy lifestyle are more inclined to participate in screening studies, a healthy screenee effect may be introduced [25]. Therefore, to generalize the ERSPC data to a general Western population, the ERSPC LE was adjusted for a potential healthy screenee effect with the World Health Organization (WHO) Dutch life tables. The relative mortality between the ERSPC and the Dutch life tables was calculated with a Poisson regression corrected for age and comorbidity [26,27]. This relative mortality was added to the LE prediction for men without csPCa.

RESULTS

For the prediction of having a biopsy-detectable csPCa, a total of 3616 men underwent sextant biopsies (PSA ≥ 3.0 ng/mL) in the first screening round ERSPC Rotterdam. A total of 313 (9%) csPCa cases were detected in addition to 572 (16%) indolent PCa cases (ISUP grade 1). Clear differences between no PCa, indolent PCa, and csPCa were noted for age, PSA, freePSA/total PSA ratio (%freePSA), prostate volume, and number of abnormal DRE/TRUS (transrectal ultrasound) findings (Table 1). Family history and IPSS did not differ between groups. The combination of PSA, age, and %freePSA (further referred to as the "basic model") was associated with a significant increase in the concordance index compared to PSA alone (0.810 versus 0.767; $p < 0.001$), and the IPA was 21%, indicating a useful prediction model with good discrimination and calibration. The addition of DRE and a rough estimate of prostate volume to the prediction model increased the concordance index even more [to 0.839 ($p < 0.001$) and 0.862 ($p < 0.001$), respectively; Table 2]. Decision curve analysis showed a positive net benefit for all models compared to the default strategy (biopsy when PSA ≥3.0 ng/mL, Figure S1). The basic model with a 5% threshold would have a net reduction of 26% (261/1000) biopsies compared to the default strategy while not increasing the missed csPCa (Table S1). The basic model was included in the final prediction tool, as it has a good balance between high predictive accuracy and practical considerations (i.e., every GP can easily use the basic model).

For men without csPCa, LE was estimated using a Weibull distribution on the ERSPC section Rotterdam (n = 19,553). The median follow-up was 15 years (interquartile range 12–17). Between 1993 and 2013, 7318 (38%) men died, 172 (2%) of whom died of PCa. The 10-year

overall survival rate was 81% (95% CI: 80–81%). A healthy screenee effect with a hazard ratio of 1.6 was found between the ERSPC screening cohort versus the general population (Table S2).

Table 1. Characteristics of 3616 men with a biopsy stratified to prostate cancer outcome from the European Randomized Study of Screening for Prostate Cancer (ERSPC).

Characteristics	No PCa *n* = 2731 (75%)	Indolent PCa (GS 3 + 3) *n* = 572 (16%)	PCa GS ≥3+ 4 *n* = 313 (9%)
Age, years, median (IQR)	66 (60–70)	67 (61–70)	68 (64–71)
PSA, ng/mL, median (IQR)	4.0 (2.5–5.7)	5.1 (3.7–7.4)	7.8 (4.8–16.0)
%FreePSA	0.22 (0.17–0.28)	0.17 (0.12–0.24)	0.12 (0.08–0.17)
Prostate volume, mL, median (IQR)	43 (33–57)	37 (29–50)	37 (29–47)
Abnormal DRE, *n* (%)	836 (31)	236 (41)	207 (66)
Abnormal TRUS, *n* (%)	795 (29)	226 (40)	208 (66)
Positive family history, *n* (%)	210 (8)	64 (11)	30 (10)
IPSS, median (IQR)	5 (2–11)	4 (1–9)	4 (1–10)

PCa= prostate cancer, GS= Gleason Score, DRE= digital rectal exam, TRUS= transrectal ultrasound, PSA= prostate-specific antigen, IQR= interquartile range, IPSS= International Prostate Symptom Score.

The estimates of LE in years for men with csPCa receiving no treatment were made on the basis of age and the Charlson comorbidity index using the SEER data (Figure 2). For men without comorbidity aged 65, 70, or 75 years, the LE was estimated to be 12.3, 10.9, and 9.7 years, respectively.

The treatment benefit for men with csPCa was estimated as a hazard ratio (HR) of 0.84 (95% CI, 0.70 to 1.01) on all-cause mortality in the PIVOT trial [19]. This overall hazard ratio (0.84) was used to estimate the absolute treatment benefit. Treatment of csPCa was expected to increase LE with 1.6 years, 1.5 years, and 1.3 years for men without comorbidity aged 65, 70, and 75 years, respectively (Figure 2). The SPCG-4's and ProtecT's HR for death by any cause were 0.74 (95% CI, 0.62–0.87) and 0.93 (95% CI, 0.65–1.35), respectively. In the sensitivity analysis, shorter LE for men with csPCa receiving treatment was found when the ProtecT's HR was used, and longer LE was found when using the SPCG-4's HR, supplement Table S3.

Table 2. Concordance index corrected for optimism and index of prediction accuracy (IPA) for individual and combined predictive performance for each variable of the risk calculator predicting prostate cancer with Gleason ≥3 + 4 in 3616 men from the ERSPC.

Univariable	Concordance index (95% CI)	IPA (%)	Multivariable	Concordance index (95% CI)	IPA (%)
PSA	0.77 (0.74–0.80)	15.4	PSA + Age	0.77 (0.74–0.80)	15.5
Age	0.59 (0.56–0.62)	0.7	PSA + %FreePSA	0.80 (0.77–0.83)	20.3
%FreePSA	0.78 (0.75–0.81)	11.1	PSA + Age + %FreePSA	0.81 (0.78–0.84)	21.0
DRE	0.67 (0.64–0.70)	3.9	PSA + Age + DRE	0.82 (0.79–0.84)	22.3
Prostate volume	0.60 (0.56–0.63)	0.7	PSA + Age + DRE + %FreePSA	0.84 (0.81–0.86)	26.3
FH	0.51 (0.49–0.52)	0.0	PSA + Age + DRE + %FreePSA + PV	0.86 (0.84–0.88)	28.3
IPSS/AUA	0.52 (0.49–0.56)	0.0	Above + TRUS	0.87 (0.85–0.90)	31.6
TRUS	0.68 (0.65–0.71)	4.5	Above + FH and IPSS/AUA	0.86 (0.84–0.88)	31.6

%FreePSA= FreePSA divided by PSA, AUA = American Urological Association symptom score, PV = prostate volume, FH = family history.

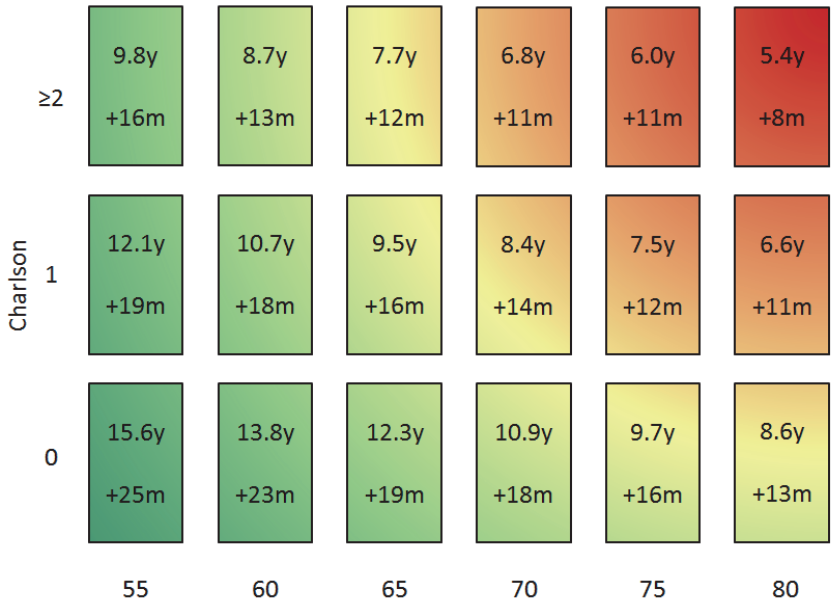


Figure 2. Life expectancy in years (y) for patients with clinically significant prostate cancer receiving no treatment based on Weibull distribution from the SEER. Secondly, gain in life expectancy (LE) in months (m) by csPCa treatment based on the relative risk reduction for all-cause mortality from the Prostate Cancer Intervention versus Observation Trial (PIVOT). The green color indicates a patient should be referred to a urologist, red indicates the patient should not be referred. Colors overlap because risk of csPCa on a current biopsy should also be weighted.

The individual risk of having csPCa, the LE, and the potential absolute treatment benefit in terms of survival rate were checked against the recommendations and are summarized in an advice for referral to secondary care in Figure 3. In this example, a 65-year-old man without comorbidity had a PSA level of 4.0 ng/mL and %freePSA of 17%. His current risk of csPCa on biopsy was 9%. The patient's life expectancy would be 12.3 years if csPCa was undetected and untreated. If the cancer was detected and treated, his life expectancy would increase by 20 months. Here, one would advise a referral for further assessment. However, a 75-year-old man with Charlson comorbidity index 2 with similar PSA and freePSA values would have a very limited absolute benefit of early detection and treatment despite a higher risk of having csPCa (15%). The latter man should not be referred to a urologist, as his potential benefit from referral would be low.

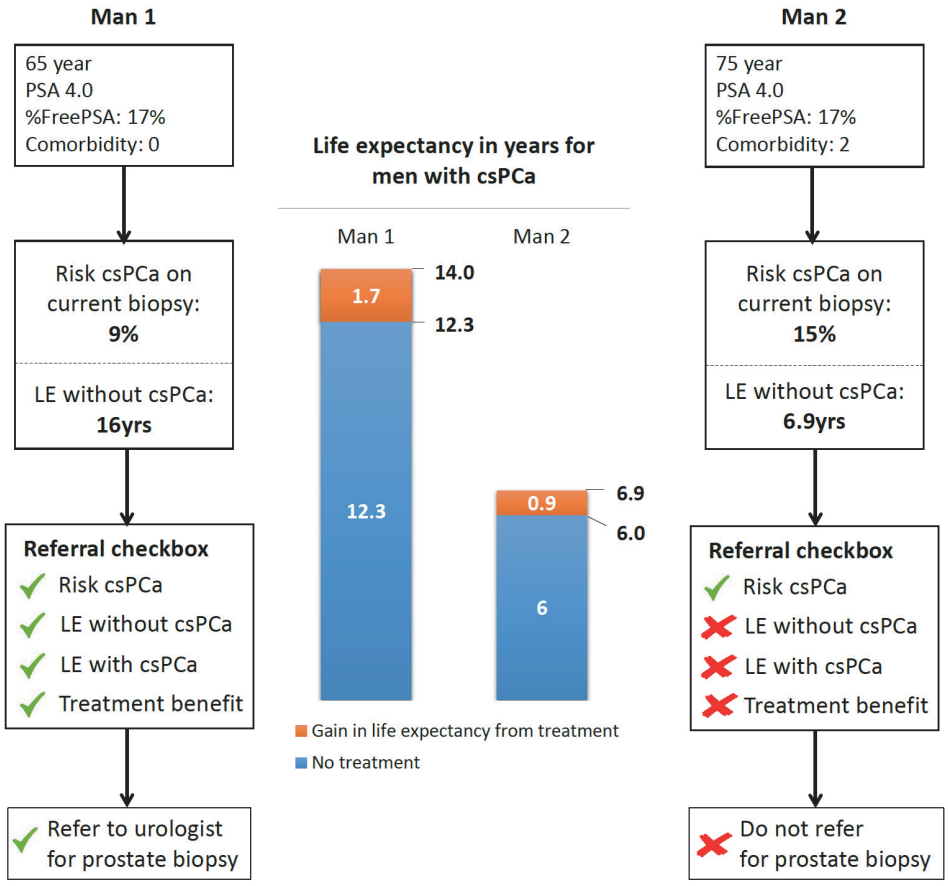


Figure 3. Output of the prediction tool for the general practitioner (GP). Displaying risk of clinically significant prostate cancer (csPCa; Gleason score ≥3 + 4) on a current biopsy, LE in years with and without csPCa, treatment benefit in years, and referral advice in two male examples.

DISCUSSION

The integrated approach described in this manuscript provides the potential gain in LE when being diagnosed and treated for csPCa. In current practice, many men are referred with a high PSA for a prostate biopsy to the urologist, while many have benign prostatic hyperplasia. Prediction tools can already reduce unnecessary referrals for biopsies [9]. However, many old men are still referred simply on the basis of having an elevated risk of having a csPCa, while it is unlikely that they will benefit from detection and treatment of their PCa. The proposed tool can help primary care physicians triage patients for timely and necessary referral for further assessment, and as such, can aid in reducing unnecessary testing, over-diagnosis, and subsequent overtreatment. This approach can thus aid in improving the unfavorable harm–benefit ratio of opportunistic PSA testing [28]. The prediction tool is easy to use, as it requires only readily accessible information and provides risk percentages supported by recommendations on how to pursue. It is suitable for Western daily clinical practice, as it has been developed on well-known, long-term, high quality cohorts, including the SEER, the PIVOT, and the ERSPC.

Prostate cancer risk calculators including patients' LEs have been published before [10,29]. However, the calculators lack recommendations and do not include treatment benefit in terms of overall survival. To estimate treatment benefit in the current prediction tool, PIVOT follow-up data were used [19]. The PIVOT data show that after nearly 20 years of follow-up surgery, localized PCa was associated with a lower all-cause or PCa-specific mortality compared to observation. Even with this long follow-up, the event rate was so low that no statistical significance was reached for the treatment effect of 0.84. The confidence interval indicated substantial uncertainty around this effect estimate (0.70 to 1.01). Other randomized clinical trials comparing PCa treatment with observation include the SPCG-4 and the ProtecT trial. The SPCG-4 with 29-year follow-up showed that surgery was associated with longer LE for men with localized PCa [20]. The ProtecT with 10-year follow-up found no clear differences between surgery, radiation, or active monitoring [21]. The mortality differences across these three studies may reflect differences in patients' characteristics, the natural history of PCa, and the difference in detection and treatment methods. Sensitivity analysis was performed using the different relative treatment effects from these studies. More treatment benefit, and thus longer LE, was predicted when the relative treatment effect from the SPCG-4 was used, and shorter LE was predicted when the relative treatment of the ProtecT was used. Unfortunately, individual treatment benefit based on patients' characteristics could not be estimated, as the numbers in all these studies are relatively small, prohibiting meaningful subgroup analysis [30]. An individual participant data meta-analysis involving the collection of the original data from the PIVOT, SPCG-4, and ProtecT would improve quality and reliability

of the treatment effect estimation. This would require collaboration between researchers and take more time and resources than extracting the results from the published reports.
The construction of our model is not without its potential weaknesses. Treatment benefit is based on 20-year-old information. Our multidimensional prediction tool needs further validation based on new screening and treatment trail data. It is important to validate the contemporary treatment effect. Improvements in prostate cancer treatment might positively affect LE. Also, the predictions are limited to the information that was available at the time of analysis. For example, we did not include other predictive factors known to affect LE, e.g., marital status, body mass index (BMI), race, and smoking. The recommendation to refer a man for further assessment is based on consensus, however, this recommendation should be seen as an aid in the shared decision-making process and not as a replacement. Treatment effect is based on overall treatment effect from the PIVOT, as no statistical differences in treatment effect between age groups or comorbidity categories were found. However, this may have been due to insufficient numbers to properly perform subgroup analysis. Although our prediction model only estimates LE, other outcome measures can influence a decision to refer for biopsy, e.g., quality of life, disease-free LE, or progression to metastatic disease. These other outcomes were not available in the datasets but should be considered when referring a patient. The SEER and the ERSPC data represent different settings (United States versus Europe). This might be a limitation, as the SEER consists of 15% African American men, while the ERSPC data mainly consist of Caucasian men. However, the SEER and the ERSPC have minimal selection bias and represent the general Western daily clinical practice. The field of prostate cancer detection is developing with imaging techniques such as mpMRI (multi-parametric magnetic resonance imaging) and PET-scans (positron emission tomography). MpMRI is known to detect more csPCa than TRUS-guided biopsies [31]. Therefore, the decision path might be improved with the inclusion of mpMRI target biopsies. However, mpMRI studies include referred men with a high suspicion of csPCa, which is represented by the high PSA and the high csPCa prevalence rate. Therefore, it is not yet well-established which patients should undergo an MRI, as the definition of “high risk of having csPCa” for initial men is not properly defined. Without a proper mpMRI screening trial with a standardized protocol, it is unobtainable to incorporate the mpMRI workflow in our model. In the future, our model should be validated for the prediction of csPCa with the inclusion of mpMRI and other novel biomarkers.

Our proposed GP prostate cancer prediction tool uses age, PSA, %freePSA, and comorbidity to provide recommendations to refer for prostate biopsy. These predictors provide a balance between predictive accuracy and practical considerations. Higher clinical impact can be achieved using a more accurate prediction on the risk of having csPCa when including DRE and prostate volume.

CONCLUSION

The estimation of life expectancy, risk of aggressive PCa, and potential benefit of prostate cancer treatment are the key aspects in the dilemma for the general practitioner and their patients regarding whether or not they should be referred for prostate biopsy. The proposed multivariable and multidimensional prediction tool needs further validation. It can provide valuable insight into the expected benefit of an early diagnosis of prostate cancer.

REFERENCES

1. Schröder, F.H.; Hugosson, J.; Roobol, M.J.; Tammela, T.L.; Zappa, M.; Nelen, V; et al. Screening and prostate cancer mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014, 384, 2027–2035.
2. Loeb, S.; Bjurlin, M.A.; Nicholson, J.; Tammela, T.L.; Penson, D.F.; Carter, H.B.; Carroll, P.; Etzioni, R. Overdiagnosis and overtreatment of prostate cancer. Eur. Urol. 2014, 65, 1046–1055.
3. Bibbins-Domingo, K.; Grossman, D.C.; Curry, S.J. The US Preventive Services Task Force 2017 Draft Recommendation Statement on Screening for Prostate Cancer: An Invitation to Review and Comment. Jama 2017, 317, 1949–1950.
4. Mottet, N.; Bellmunt, J.; Bolla, M.; Briers, E.; Cumberbatch, M.G.; De Santis, M.; Fossati, N.; Gross, T.; Henry, A.M.; Joniau, S.; et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. Eur. Urol. 2017, 71, 618–629.
5. Cooperberg, M.R. Implications of the new AUA guidelines on prostate cancer detection in the U.S. Curr. Urol. Rep. 2014, 15, 420.
6. Ankerst, D.P.; Hoefler, J.; Bock, S.; Goodman, P.J.; Vickers, A.; Hernandez, J.; Sokoll, L.J.; Sanda, M.G.; Wei, J.T.; Leach, R.J.; et al. Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer. Urology 2014, 83, 1362–1367.
7. Roobol, M.J.; Steyerberg, E.W.; Kranse, R.; Wolters, T.; van den Bergh, R.C.; Bangma, C.H.; Schröder, F.H. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. Eur. Urol. 2010, 57, 79–85.
8. Nam, R.K.; Toi, A.; Klotz, L.H.; Trachtenberg, J.; Jewett, M.A.; Appu, S.; Loblaw, D.A.; Sugar, L.; Narod, S.A.; Kattan, M.W. Assessing individual risk for prostate cancer. J. Clin. Oncol. 2007, 25, 3582–3588.
9. Ries, L.A.G.; Melbert, D.; Krapcho, M.; Mariotto, A.; Miller, B.A.; Feuer, E.J.; Clegg, L.; Horner, M.J.; Howlader, N.; Eisner, M.P.; et al. (Eds.) SEER Cancer Statistics Review, 1975–2004; National Cancer Institute: Bethesda, MD, USA, 2007. Available online: seer.cancer.gov/csr/1975_2004/ (accessed on 16 October 2009).
10. Nam, R.K.; Satkunavisam, R.; Chin, J.L.; Izawa, J.; Trachtenberg, J.; Rendon, R.; Bell, D.; Singal, R.; Sherman, C.; Sugar, L.; et al. Next-generation prostate cancer risk calculator for primary care physicians. Can. Urol. Assoc. J. = Journal de L'Association des Urologues du Canada 2018, 12, e64–e70.
11. Roobol, M.J.; Schröder, F.H.; Hugosson, J.; Jones, J.S.; Kattan, M.W.; Klein, E.A.; Hamdy, F.; Neal, D.; Donovan, J.; Parekh, D.J.; et al. Importance of prostate volume in the European Randomized Study of Screening for Prostate Cancer (ERSPC) risk calculators: Results from the prostate biopsy collaborative group. World J. Urol. 2012, 30, 149–155.
12. Kattan, M.W.; Gerds, T.A. The index of prediction accuracy: An intuitive measure useful for evaluating risk prediction models. Diagn. Progn. Res. 2018, 2, 7.
13. Van Calster, B.; Wynants, L.; Verbeek, J.F.; Verbakel, J.Y.; Christodoulou, E.; Vickers, A.J.; Roobol, M.J.; Steyerberg, E.W. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur. Urol. 2018, 74, 796–804.
14. Albertsen, P.C.; Fryback, D.G.; Storer, B.E.; Kolon, T.F.; Fine, J. The impact of co-morbidity on life expectancy among men with localized prostate cancer. J. Urol. 1996, 156, 127–132.
15. Albertsen, P.C.; Moore, D.F.; Shih, W.; Lin, Y.; Li, H.; Lu-Yao, G.L. Impact of comorbidity on survival among men with localized prostate cancer. J. Clin. Oncol. 2011, 29, 1335–1341.

16. Albertsen, P.C.; Hanley, J.A.; Fine, J. 20-year outcomes following conservative management of clinically localized prostate cancer. JAMA 2005, 293, 2095–2101.

17. Roobol, M.J.; Kranse, R.; Bangma, C.H.; van Leenders, A.G.; Blijenberg, B.G.; van Schaik, R.H.; Kirkels, W.J.; Otto, S.J.; van der Kwast, T.H.; de Koning, H.J.; et al. Screening for prostate cancer: Results of the Rotterdam section of the European randomized study of screening for prostate cancer. Eur. Urol. 2013, 64, 530–539.

18. de Carvalho, T.M.; Heijnsdijk, E.A.; de Koning, H.J. Estimating the risks and benefits of active surveillance protocols for prostate cancer: A microsimulation study. BJU Int. 2017, 119, 560–566.

19. Wilt, T.J.; Jones, K.M.; Barry, M.J.; Andriole, G.L.; Culkin, D.; Wheeler, T.; Aronson, W.J.; Brawer, M.K. Follow-up of Prostatectomy versus Observation for Early Prostate Cancer. N. Engl. J. Med. 2017, 377, 132–142.

20. Bill-Axelson, A.; Holmberg, L.; Garmo, H.; Taari, K.; Busch, C.; Nordling, S.; Häggman, M.; Andersson, S.O.; Andrén, O.; Steineck, G.; et al. Radical Prostatectomy or Watchful Waiting in Prostate Cancer—29-Year Follow-up. New Engl. J. Med. 2018, 379, 2319–2329.

21. Hamdy, F.C.; Donovan, J.L.; Lane, J.A.; Mason, M.; Metcalfe, C.; Holding, P.; Davis, M.; Peters, T.J.; Turner, E.L.; Martin, R.M.; et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N. Engl. J. Med. 2016, 375, 1415–1424.

22. Schröder, F.H.; Hugosson, J.; Roobol, M.J.; Tammela, T.L.; Ciatto, S.; Nelen, V.; Kwiatkowski, M.; Lujan, M.; Lilja, H.; Zappa, M.; et al. Screening and prostate-cancer mortality in a randomized European study. N. Engl. J. Med. 2009, 360, 1320–1328.

23. Fitzmaurice, C.; Dicker, D.; Pain, A.; Hamavid, H.; Moradi-Lakeh, M.; MacIntyre, M.F.; Allen, C.; Hansen, G.; Woodbrook, R.; Wolfe, C.; et al. The Global Burden of Cancer 2013. JAMA Oncol. 2015, 1, 505–527.

24. Buuren, S.V.; Groothuis-Oudshoorn, K. Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 2011, 45, 1–67.

25. Ilic, D.; Neuberger, M.M.; Djulbegovic, M.; Dahm, P. Screening for prostate cancer. Cochrane Database Syst Rev. 2013, (1): CD004720.

26. Dauxois, J.Y.; Guilloux, A.; Kirmani, S.N. Estimation in a competing risks proportional hazards model under length-biased sampling with censoring. Lifetime Data Anal. 2014, 20, 276–302.

27. Centraal Bureau voor de Statistiek (CBS) (2016): Gezondheidsenquête 2015—GECON 2015. DANS. Available online: doi.org/10.17026/dans-xwr-m26w (accessed on 20 September 2018).

28. Hugosson, J.; Godtman, R.A.; Carlsson, S.V.; Aus, G.; Grenabo Bergdahl, A.; Lodding, P.; Pihl, C.G.; Stranne, J.; Holmberg, E.; Lilja, H. Eighteen-year follow-up of the Goteborg Randomized Population-based Prostate Cancer Screening Trial: Effect of sociodemographic variables on participation, prostate cancer incidence and mortality. Scand. J. Urol. 2018, 52, 27–37.

29. Kent, M.; Penson, D.F.; Albertsen, P.C.; Goodman, M.; Hamilton, A.S.; Stanford, J.L.; Stroup, A.M.; Ehdaie, B.; Scardino, P.T.; Vickers, A.J. Successful external validation of a model to predict other cause mortality in localized prostate cancer. BMC Med. 2016, 14, 25.

30. Kent, D.M.; Steyerberg, E.; van Klaveren, D. Personalized evidence-based medicine: Predictive approaches to heterogeneous treatment effects. BMJ 2018, 363, k4245.

31. Kasivisvanathan, V.; Rannikko, A.S.; Borghi, M.; Panebianco, V.; Mynderse, L.A.; Vaarala, M.H.; Briganti, A.; Budäus, L.; Hellawell, G.; Hindley, R.G.; et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. N. Engl. J. Med. 2018, 378, 1767–1777.

SUPPLEMENTARY MATERIALS

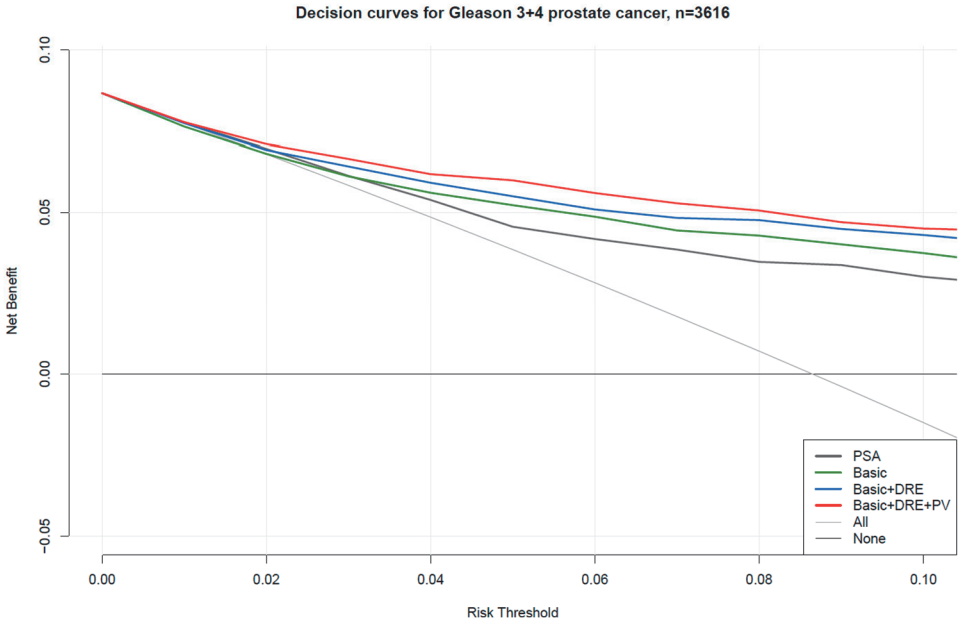


Figure S1. Decision curve analysis for different prediction models incorporating prostate measurements in 3616 patients from ERSPC.

Table S1. Clinical impact of the three prostate cancer risk calculators.

Net benefit of the three prostate cancer prediction models for the general practitioner																
Threshold	**Basic model: PSA+FreePSA+Age**					**Basic + DRE**					**Basic+DRE+PV(DRE)**					**Ref. strategy**
	A	B	C	ΔD	E	A	B	C	ΔD	E	A	B	C	ΔD	E	D
Total, n		87 (0)	158 (0)				87 (0)	158 (0)				87 (0)	158 (0)			
3%	320	7 (8)	27 (17)	3	90	428	7 (8)	41 (26)	6	188	466	6 (7)	48 (31)	8	263	58
5%	**526**	**13 (15)**	**57 (36)**	**14**	**261**	**606**	**15 (17)**	**70 (44)**	**17**	**313**	**622**	**11 (13)**	**67 (43)**	**22**	**406**	**38**
10%	770	30 (34)	101 (64)	52	471	787	27 (31)	102 (65)	58	521	791	25 (29)	97 (61)	60	539	-15

A: Absolute number of biopsy reduction per 1000

B: Absolute number of missed PCa GS ≥3+4 per 1000 (% of total PCa GS ≥3+4)

C: Absolute number of missed indolent PCa per 1000 (% missed of total indolent PCa)

D: Net benefit x 1000; ΔD: Additional net true positives per 1000 men compared with reference strategy

E: Net reduction of biopsies per 1000 men, compared to the reference strategy (biopsy is advised when PSA >3.0ng/ml); formula: ΔNB/(threshold/(1-threshold))*1000.

Table S2. Characteristics of the ERSCP screenings arm vs. 1458 men from the Dutch general population (year 2015).

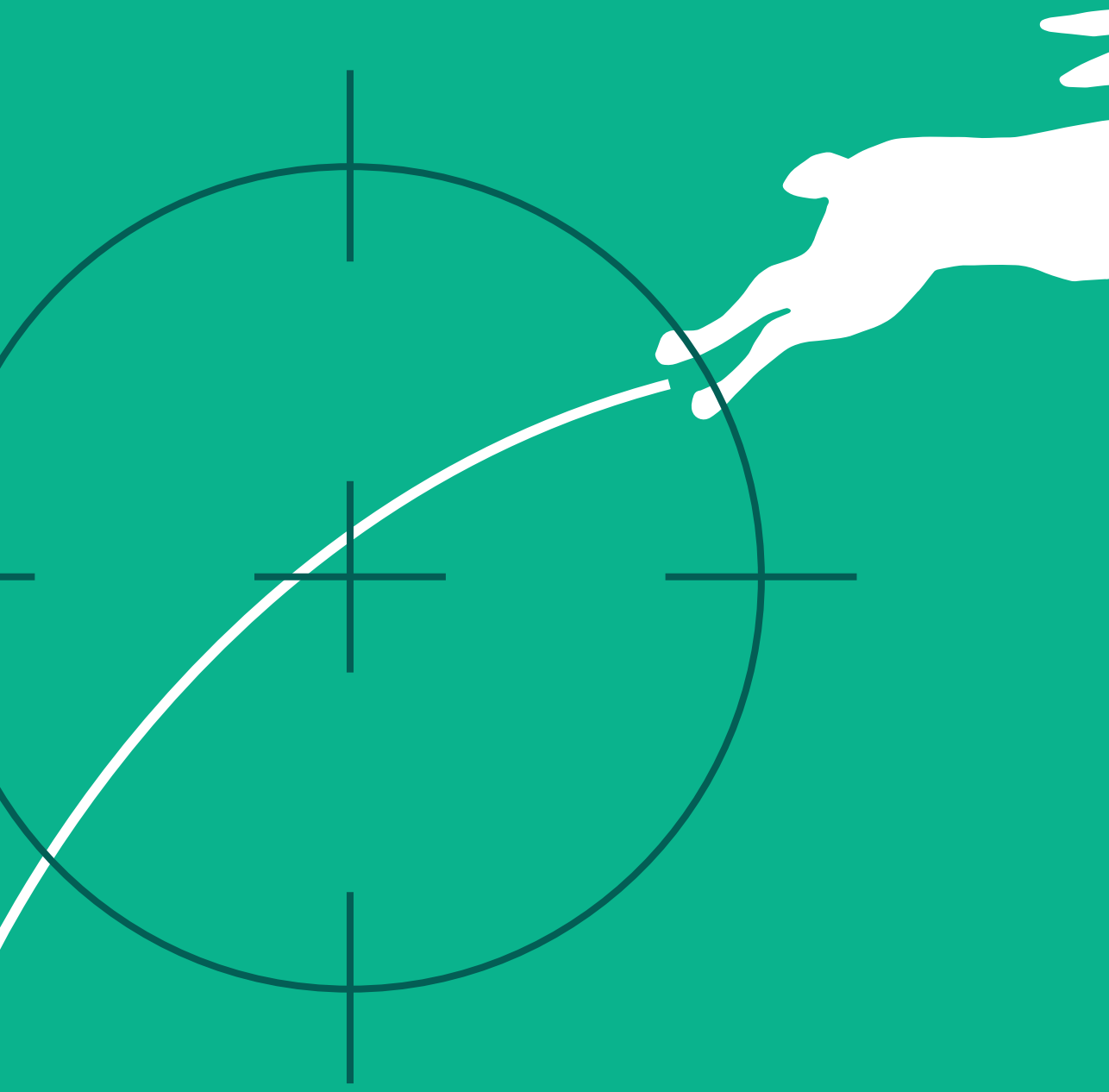
	ERSPC n=19247	**General Dutch Population n=1458**
Age, median (IQR)	63 (59-68)	66 (60-72)
Charlson comorbidity index		
0	14095 (73)	801 (55)
1	4332 (23)	457 (31)
2	662 (3)	200 (14)
missing	158 (1)	-

Table S3. Sensitivity analysis of the life expectancy of men with clinically significant prostate cancer receiving prostate cancer treatment using the different treatment hazard ratios in the prostate cancer treatment trials.

Life expectancy in years Charlson comorbidity score = 0				
Age	Treatment - Hazard rate (95% CI)			
	None	ProtecT 0.93 (0.65 – 1.35)	PIVOT 0.84 (0.70 - 1.01)	SPCG-4 0.74 (0.62 – 0.87)
55	15.6	16.3	17.7	19.4
65	12.3	12.8	13.9	15.3
75	9.7	10.1	11.0	12.0

Life expectancy in years Charlson comorbidity score = 1				
Age	Treatment - Hazard rate (95% CI)			
	None	ProtecT 0.93 (0.65 – 1.35)	PIVOT 0.84 (0.70 - 1.01)	SPCG-4 0.74 (0.62 – 0.87)
55	12.1	12.6	13.7	15.0
65	9.5	9.9	10.8	11.8
75	7.5	7.8	8.5	9.3

Life expectancy in years Charlson comorbidity score ≥ 2				
Age	Treatment - Hazard rate (95% CI)			
	None	ProtecT 0.93 (0.65 – 1.35)	PIVOT 0.84 (0.70 - 1.01)	SPCG-4 0.74 (0.62 – 0.87)
55	9.8	10.2	11.1	12.1
65	7.7	8.0	8.7	9.5
75	6.0	6.3	6.9	7.5



CHAPTER 8

Discussion and future recommendations
Summary
Summary in Dutch

DISCUSSION AND FUTURE RECOMMENDATIONS

The aim of this thesis was to contribute to the development, evaluation and implementation of personalized-based medicine with the use of prediction models in the field of early prostate cancer (PCa) detection. Personalized-based medicine can provide higher clinical utility as it uses an absolute risk estimate compared to the relative risk estimated with classic evidence-based medicine. For example, in the European Randomized Study of Screening for Prostate Cancer (ERSPC), the 16-year cumulative risk of PCa death per 1000 men in the screening group was 7.17, versus 8.92 in the control group. This means that on a relative scale, the screening benefit is 7.17 / 8.92 = 0.80 or, stated otherwise, the relative PCa mortality reduction was 20% [1]. On an additive scale, however, the absolute risk reduction or screening benefit is 8.92 per 1000 minus 7.17 per 1000 = 1.75 per 1000. This figure of 1.75 per 1000 men screened is equivalent to 1 per 571 men (1000/1.75). So, when translating relative to absolute numbers, it meant that to avoid one man dying from PCa, 571 men had to be (repeatedly) screened. This 1-to-571 ratio also indicates that the majority of men participating in screening programs are at risk of undergoing unnecessary tests with related potentially harmful side-effects ranging from worry about a test result, confrontation with the diagnosis of cancer that would never cause harm if undetected, to lethal sepsis as a complication after prostate biopsy. According to the 16-year follow-up of the ERSPC trial, the number of cases needed to diagnose for averting one PCa death was 18. Unfortunately, it is unknown who will benefit or harmed.

In the first part of this thesis, the one-size-fits-all approach based on classic evidence-based medicine (i.e. formulating guidelines for future patients based on the average outcome of current patients) was described for PCa screening and AS, and the importance of risk-stratification. The second part described how multivariable PCa prediction models can formulate an individual approach for a current patient on the basis of outcomes of comparable patients in the past. Prediction models can help to reduce unnecessary testing and, as such, be of aid to both patients and physicians by improving informed decision-making.

CLASSIC EVIDENCE-BASED MEDICINE: OUTCOMES OF PROSTATE CANCER SCREENING AND ACTIVE SURVEILLANCE

Prostate cancer screening

The ERSPC was set up in the 1990s and is the largest randomized PCa screening trial [2]. In this classic evidence-based medicine project with its one-size-fits-all approach, only 25% of

all biopsies confirmed the suspicion of PCa with the use of PSA test [3]. This implies that 75% of men at that point in time were unnecessarily biopsied and exposed to the negative consequences of attending an early detection program; perhaps the positive news was that no PCa was found [4]. Obviously, in view of a balanced harm-benefit ratio, there has been and still is a focus on finding ways to reduce the number of these potentially unnecessary biopsies. This classic one-size-fits all screening approach not only results in many unnecessary biopsies, it also leads to intensive retesting procedures as both indication for biopsy and the biopsy procedure itself do not exclude a future PCa diagnosis.

But first, to gain more insights in the potential benefit of additional tests and repeating biopsy procedures, the long-term follow-up data of a one-size-fits all screening algorithm were studied. PCa incidence data of all men with a PSA below the general advised threshold of 3.0 ng/ml for referral for biopsy and those with a negative (i.e. benign result) prostate biopsy was collected. These data provide an insight into the so-called false negative (FN) test results and were reported in Chapter 2 [5]. After data analysis, it became clear that the FN rates for men with a PSA $<$3.0 ng/mL and those who had a sextant Bx with benign result were, although not negligible, extremely low. So, despite the one-size-fits-all approach almost detecting all clinically significant PCa, this raises the question whether personalized-medicine should be introduced to diagnose these low numbers of FN and, if so, at what cost? In this study, we used sextant biopsies which are known for their poor detection rate [6], but nevertheless the FN and the PCa mortality were low. Almost all PCa were detected by screening with PSA and performing prostate biopsy with sextant biopsies. This means that over the course of 15 years, the maximum achievable increase in detection of potentially life-threatening PCa by applying additional diagnostic tools in the screening algorithm like novel biomarkers and multiparametric Magnetic Resonance Imaging (mpMRI) might be limited in terms of detecting more harmful PCa.

These additional tools should nevertheless be considered within the broader context of the PSA-screening debate as they can reduce the number of unnecessary biopsies and the considerable overdiagnosis of indolent PCa [7]. A Cochrane review suggest that PCa diagnosis with mpMRI, which includes mpMRI with or without mpMRI-targeted biopsy, is superior to systematic biopsies and reduces unnecessary biopsies and detection of indolent PCa [8, 9]. However, the use of mpMRI still misses some clinically significant PCa, and it is unclear how MRI performs in a screening setting [10]. The adoption of proper stratification for high and low-risk PCa (before application of additional diagnostic tools including targeted mpMRI biopsy), will help to create an improved balance between harms and benefits of PCa screening and to make a further transition from the classic evidence-based medicine to personalized medicine. Moreover, proper risk stratification with applicable adequate imaging and biopsy procedures at the first screening exam will also result in recommendations

regarding when to refrain from further testing, and/or for longer retest intervals if results are benign.

It is important to know that the one-size-fits all approach will only provide a relative risk reduction for those individuals who have exceeded the decision threshold. However, an individual close to the decision threshold (PSA 2.9 ng/ml) might have a higher risk reduction than an individual far from reaching the threshold (PSA 0.1 ng/ml); in the one-size-fits-all method both men would receive the same treatment and would not receive a biopsy. The PCa screening one-size-fits-all with PSA had good results, especially in men with PSA <1.0 ng/ml where almost no PCa was found [11]. Therefore, from an absolute risk perspective, the PSA <1.0 ng/ml threshold would be reason enough not to propose further actions. However, if a cheap 'magic marker' would find those few PCa cases, it could be beneficial to screen these men.

The one-size-fits-all is reasonable for a relative risk reduction, but the key is to translate this relative risk to an absolute risk reduction. This absolute risk reduction should preferably be estimated on the basis of an individual risk profile instead of the possible reduced average of the (sub)group. This risk profile can be formulated with prediction models combining patient characteristics and test results such as mpMRI findings and biomarkers. Nonetheless, the shift towards personalized medicine with prediction models is limited by the fact that not everything can be observed and thus not be incorporated in models. For example, why are there still men with aggressive PCa despite very low PSA levels and why are there men with high PSA values that do not have PCa? Obviously, there is more than the PSA value that determines the presence, more relevant information can give more insights, but a considerable part of the causes underlying the development of PCa are still unknown. These causes can even be so complex that it can be seen as chaos. Table 1 and Figure 1 summarizes the clinical benefit of the one-size-fit approach and the personalized medicine approach. The personalized medicine approach has more benefit compared to the one-size-fits-all approach in general, as, using additional predictors, it can improve the distinctiveness of men around the risk threshold. Also, it can more adequately estimate the expected benefit of treatment for a specific individual as the absolute benefit depends on the underlying absolute risk. Thus, instead of using a relative risk of the entire group, it estimates the absolute risk. If the absolute risk is low, low absolute benefit might be expected, and if absolute risk is high, more absolute benefit can be gained. A prediction model improves the categorizing of low risk and high risk more efficiently, thus more clinical utility can be gained.

Table 1. Comparison of classic evidence-based medicine with personalized medicine.

	Classic evidence-based medicine	Personalized medicine
Risk estimate	relative	absolute
Outcome	odds ratio, relative risk	risk percentage of outcome
Study design	randomized trial or cohort study (small is enough)	large randomized trial (phase ≥III trials)/large cohort study with aggregation of overall results using prediction model
Reference class	wide	narrow
Decision-making	population-based decision-making	patient-specific decision-making
Clinical utility	cost-effectiveness at group level	higher cost-effectiveness can be achieved when risks are used in the shared decision-making process and when they are implemented guidelines

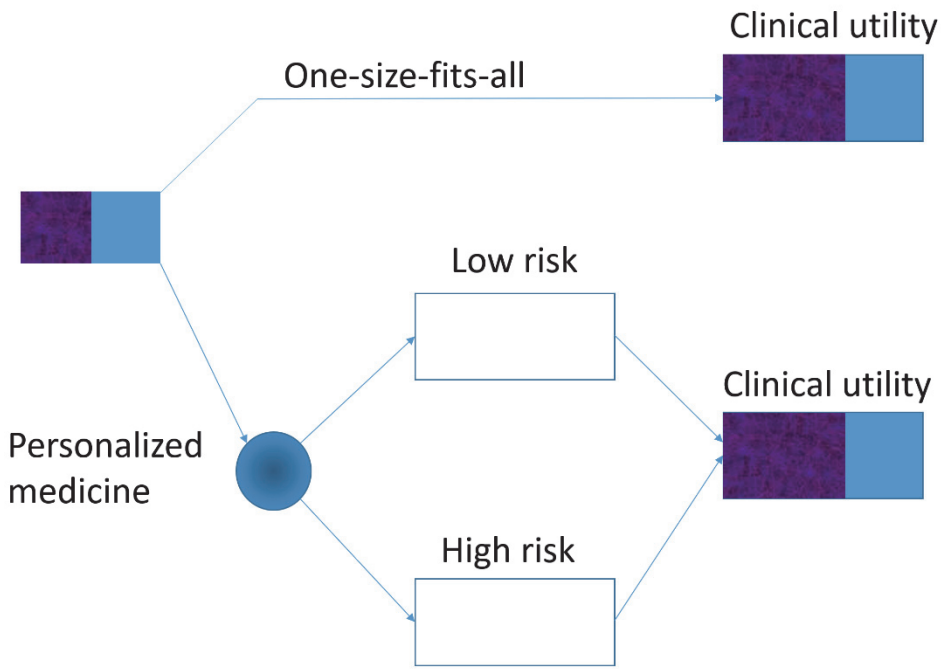


Figure 1. Clinical utility of a one-size-fits-all approach vs. personalized medicine, dark purple color indicates the achievable clinical utility in for example in terms of cost effectiveness.

The 15-year timeframe of screening with PSA and sextant prostate biopsy show that the absolute PCa mortality among the screened men was 0.5% [1], however both tests are known for their poor accuracy [12]. Still PCa mortality by PSA screening is considerably lower compared to the cumulative PCa mortality in the Dutch general population (3–5%), and the risk of dying due to PCa according to the Surveillance, Epidemiology, and End Results (SEER)

data, showing risks of 2.6%, 2.8% and 2.9% for men aged 50, 60 and 70 respectively [13]. Note that almost half of the men who died from PCa in the ERSPC trial with a previous negative Bx were not compliant with the follow-up scheme, and PCa mortality might be lower with adequate compliance [14]. To improve compliance, widening the interval of PCa screening is recommended [15].

Active surveillance

Active surveillance (AS) is the key management strategy for low-risk, localized PCa [16]. It is considered safe, as PCa mortality is comparable with active treatment (AT) when applied to a similar low PCa risk group [17] [18]. Accurate selection of men with low-risk disease and subsequent follow up is vital to AS success. However, it remains unclear whether current evidence-based guidelines for AS recruitment and ongoing surveillance are congruent with practice. Chapter 3 describes the adherence to site-specific AS inclusion criteria and follow-up biopsy schedules as assessed in the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance consortium (GAP3), comprising 25 different patient cohorts.

The adherence to inclusion criteria was good or excellent across most AS programs, although slight non-adherence was observed for PSA and Gleason Score (GS) on diagnosis. Recent evidence has emerged that this may be due to extending the criteria from only including men with GS 3+3 to including those with Gleason Score 3+4, without formally expanding the inclusion criteria. Evidence for the use of AS in men with intermediate-risk disease remains equivocal. Some prospective cohorts have demonstrated results supporting AS in well-selected GS 3+4 disease [19], although these remain inferior to men with GS 3+3 tumors [20]. Upfront risk stratification of these men into favorable and unfavorable intermediate-risk disease is required [21]. Secondary growth patterns can differentiate the more aggressive PCa in the GS 3+4 group (as discussed in Chapter 5). The prospective evaluation of their safety in AS, to justify widespread broadening of inclusion criteria should be studied in this context.

In terms of adhering to the follow-up schedule with repeat biopsy, in all centers we found that adherence was high at first, but that it decreased with time. This emphasizes the importance of accurate risk stratification, and the option of evidence-based follow-up intervals for AS: longer intervals for low-risk men. In our study, no difference in compliance between an annual biopsy schedule or longer biopsy intervals was found. However, repeat biopsy was commonly performed later than in the scheduled time with high biopsy frequency, and earlier in protocols with longer between-biopsy intervals. Therefore, existing schedules may have to be altered to define an optimal middle ground. This can be done in a personalized evidence-based manner. Simulation studies suggest the reduction of repeat biopsies is feasible, with only

minor delays in the detection of disease progression and an increase of PCa-specific mortality [22].

Diminishing adherence to repeat biopsies may be either patient-driven or clinician-driven. The effect of biopsy morbidity and psychological distress on adherence to an AS treatment strategy is unclear [23]. It is also unclear whether frequency of follow up and non-adherence has an impact on oncological or survival outcomes. Investigating the clinical impact of poor biopsy adherence is recommended, where it is important to unravel the personal risk of treatment against the expected discomfort associated with a repeat biopsy. This will then be the basis of a dynamic and personalized risk-based AS follow-up protocol where cancer control is secured with minimal burden to the patient.

Current research on mpMRI in combination with targeted biopsies is only reported if an abnormality is seen; this could be a way to minimize burden and, as such, increase adherence to the proposed examinations. This however implies that a (considerable) number of men in AS will have to be followed, only on the basis of mpMRI. Whether or not this has an impact on disease-specific mortality is unknown. In addition, targeted mpMRI-guided biopsy detects more clinically significant disease than TRUS biopsy, but misses significant cancers found on systematic transperineal template biopsy [8, 9]. Beyond mpMRI, which obviously needs more research in the AS setting, emerging tools such as PSMA PET imaging and genomic biomarkers may also provide further, individualized reassurance of the safety of continuing surveillance in selected patients, whilst avoiding repeat biopsies.

TOWARDS PERSONALIZED MEDICINE

Prediction models

The predictive performance of a prediction model is conventionally evaluated using statistical properties such as discrimination and calibration [24-26]. Discrimination measures whether a prediction model is able to discriminate between, for example, men with and without PCa. This can be calculated with the area under the receiver operating characteristic (ROC) graph or c-index. The ROC curve displays the sensitivity of the model specifications as a function of 1 minus specificity or true positive rates vs. false positive rates. The area below the ROC curve (AUC) is derived and presented as a number between 0 and 1. AUCs near the 0.5 value mean poor discriminative power of the model, AUC values approximating 1.0 indicate excellence performance.

Calibration refers to the agreement between observed and predicted outcomes. If a model predicts a 10% risk of having PCa on biopsy, then the observed proportion of PCa should be approximately 10 out of 100 patients. The calibration of a model is usually tested with a calibration plot to display the fit for the entire range of predicted probabilities – from 0% to 100%. However, discrimination and calibration do not provide sufficient insights into what the

model's added value is compared with standard health care strategy. For example, how much increase in area under the ROC is compatible with what additional benefit? Moreover, is calibration or discrimination more important in improving the predictive accuracy of a prediction model? Net Benefit (NB) and Decision Curve Analysis (DCA) are used to analyze the added value in terms of clinical utility. These concepts also take the underlying probability of having the condition to be detected into account. The first paper on DCA by Vickers et al. (2006) has been cited over 1000 times [27]. NB and DCA are increasingly used, and it has been proposed that they should be implemented in the "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" (TRIPOD) statement [28]. Still interpretation of NB is difficult as it is an absolute number; it combines the number of true positives and false positives into one single number by weighing the question: "How many biopsies are you willing to perform in order to find one csPCa". To form a DCA, the NB is displayed on the y-axis along risk thresholds of underlying disease probabilities on the x-axis, see Chapter 4 [29]. The absolute value of NB depends on the absolute risk [10]. So, if the a-priori likelihood of PCa is low, then the gain by an additional diagnostic test is small. Instead of interpreting the NB on its own, it should be addressed in comparison with the default strategy or a strategy where no intervention has been done. Once the NB of a prediction model at a clinically relevant risk threshold is higher than the NB of the default strategy, the prediction model is said to be clinically useful. The relative added value of the prediction model can then be calculated within this comparison. This raises the question of how the threshold should be chosen, as each individual might answer the question, "How many biopsies are you willing to perform in order to find one csPCa?" differently. Still if the range of acceptable thresholds provides a higher NB in comparison with the alternative, the prediction model is useful. It is important to realize is that NB and DCA are most commonly performed in the same dataset; thus, an external validation is still required [10]. Miscalibration of the model can significantly change the clinical utility of a prediction model [30].

Updating prediction models

As stated earlier, population-wide screening with PSA reduces PCa mortality, but with the downside of overdiagnosis and overtreatment [1, 31]. Men considering a PSA test would like to be informed about their benefits, possibilities of false positive and false negative results, and the extent of overdiagnosis and overtreatment [32]. However, candidates for prostate biopsy are only willing to accept a small increase of their chance of dying from PCa when considering a reduction of their chances of possible overdiagnosis and overtreatment [33]. The ERSPC risk calculator number 3 (RC3) is able to reduce unnecessary biopsy taking by 30%, while still detecting the majority of clinically significant PCa [34]. The RC3 calculator uses pre-biopsy information on PSA, digital rectal examination (DRE), and prostate volume, to predict the probability of a biopsy-detectable PCa, and, more specifically, potentially aggressive PCa defined as Gleason Score (GS) $\geq 3 + 4$ (ISUP grade ≥ 2) [35]. Nevertheless,

the predictive accuracy is not 100%, thus the prediction model requires continuous updates and adaptations [36]. In Chapters 5 and 6 report that the ERSPC-RC was improved by incorporating contemporary pathologic biopsy classifications that better reflect disease burden, and by the novel biomarker 4Kscore.
The clinical course of ISUP grade ≥2 PCa is heterogeneous. Kweldam et al. found that the presence of a cribriform pattern (CR) and intraductal carcinoma (IDC) in ISUP grade 2 radical prostatectomy specimens were major predictive factors for the occurrence of distant metastases; these factors were also related to PCa-specific death [37]. These observations on growth pattern were confirmed when studying biopsy specimen [38]. More accurate disease classification based on biopsy specimen is therefore indispensable for treatment decision-making. The ERSPC-RC was updated for these new pathological insights (Figure 2). When using the ERSPC-Cribriform-RC, a urologist can substantially reduce the number of patients undergoing biopsy while minimizing overdiagnosis. The improved ERSPC-Cribriform-RC diminishes overdiagnosis by 34% at a 2% missed diagnosis of clinically relevant PCa, which is a considerable improvement compared to the original ERSPC-RC. The implication of incorporating this pathological finding is significant as the largest AS cohorts by Klotz et al. showed that after a long follow-up, metastasis occurred more often in men with ISUP grade ≥2 disease at primary diagnosis [39]. Knowing that more than half of the men with ISUP grade ≥2 PCa in our study cohort had CR/IDC+ status, it is plausible that the patients in the Klotz et al. study had similar characteristics that explain the rate of observed metastases. Furthermore, in the ProtecT study, a higher number of men in the active monitoring arm suffered metastatic disease [17]. Unfortunately, the prognostic value of CR/IDC status was not studied.

Besides updating the RC with pathohistological findings, other findings like patient characteristics (age) and biomarkers can be used to further improve the prediction model. Currently, many commercial biomarkers are offered by companies with active marketing strategies, pointing towards further improvement of risk-based patient selection for prostate biopsy. One of these novel biomarkers is the four-kallikrein panel, which is combined with age and DRE outcome into a prediction tool called the 4Kscore [40]. To assess the clinical utility of the 4Kscore, it is necessary to independently compare it with the ERSPC-Cribriform-RC. Both RCs in the same development dataset were analyzed, as no other independent cohort was available. Both ERSPC-Cribriform-RC and the 4Kscore appeared to have similar predictive performance and clinical utility. In Chapter 6, a reduction of two-thirds of all biopsies can be achieved when compared to biopsies taken solely based on PSA. Adding new predictors to prediction models remains a tradeoff when, in addition to clinical utility, aspects such as patient burden and costs are important considerations. Perhaps it is human nature to assume that novel (and often more expensive) biomarkers or models are better than the existing simpler risk prediction tools. However, it is crucial that new developments like MRI or

proteomics and/or genomics-based devices are extensively compared to existing methods to ensure that they can provide more relevant pre-biopsy information whilst not overlooking aspects like resources and availability. Preferably, these new developments should be studied in the context of a new screening setting as the data from the ERSPC are more than two decades old, are predominantly of Caucasian men, and with sextant biopsies being applied.

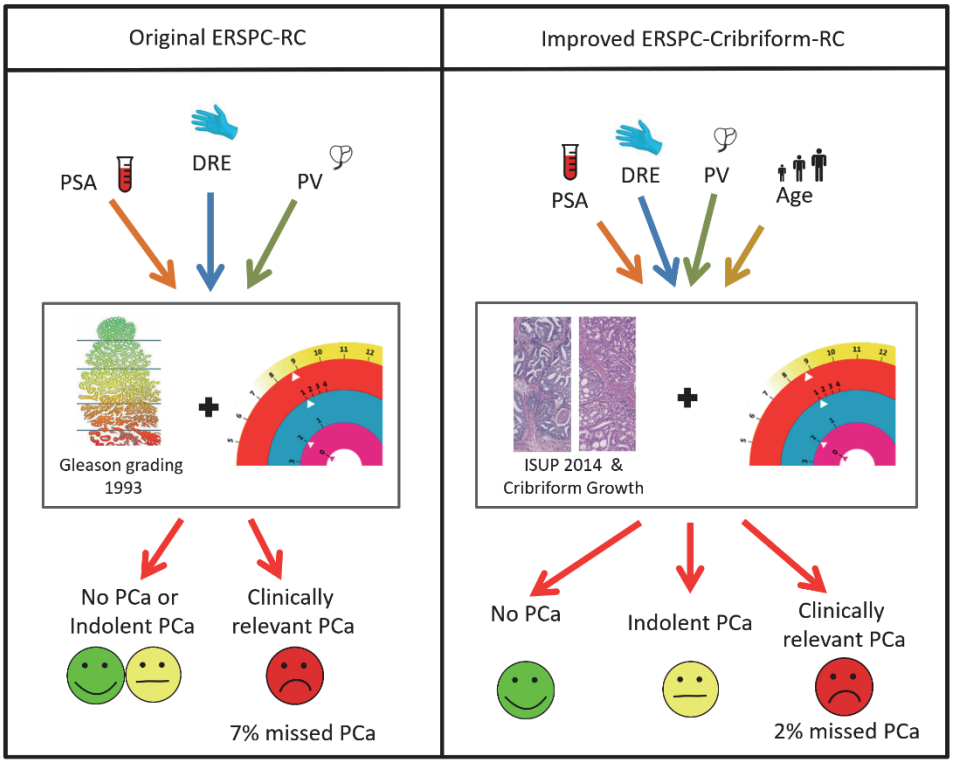


Figure 2. The original risk calculator in comparison with the updated ERSPC-Cribriform-RC including cribriform pathological findings. The improved RC distinguishes clinically relevant PCa better and reduces the number of missed diagnoses of clinically relevant PCa. PSA: Prostate Specific Antigen, DRE: Digital Rectal Exam, PV: Prostate Volume measured by DRE. ISUP: International Society of Urological Pathology. Figure published in UroToday / Gu OncToday

Prediction models in clinical practice

Many prediction models have been developed, but only a few will eventually be used in daily practice to aid shared decision-making. Several conditions have to be met before a prediction model can be used successfully in daily practice. First, the prediction model needs to show real clinical utility, that is evidence-based benefit when using the prediction model over the default strategy [41]. It should therefore be externally validated and, if needed, updated to the

new setting to improve clinical utility [10, 42]. Second, the model should be easy to use and inexpensive when applied in the screening setting. Optimizing the model's user-friendliness can be achieved by providing the outcome automatically once data are available in the electronic patient file, instead of filling in an independent form which is time-consuming for both the doctor and patient. Third, the model's outcome and the recommendations should be straightforward and understandable for both doctor and patient [43, 44]. The latter may be more complex than expected because of the concept of risk prediction itself, where the outcome is not a certainty for each specific case, but more a possibility of a certain outcome. There is no certainty that the patient has the disease. Although many people might think in these terms, they should be educated with risk management before using a prediction model; it is just a tool that aids the decision-making process, just like a scalpel is a tool used by a surgeon to operate. Finding the right scalpel without knowing how to operate, for example, is absurd.

In the same context, prediction models may be misused if there is a lack of understanding of risk management. Therefore, a risk threshold to undertake specific actions (i.e. performing a prostate biopsy) should be discussed before even applying a prediction model. To educate people about prediction models, their outcomes must be explained [45]. For example, if the risk of having a clinically significant PCa is 20%: this percentage or probability means that for every 100 patients with the same characteristics and test results, we expect that 20 men will be diagnosed with the cancer, but 80 either will have no PCa or will be diagnosed with indolent PCa which rarely causes health problems. However, this 20% risk outcome combined with the explanation may still not be sufficient, and a recommendation is still needed. However, that recommendation comes from a cut-off. An analogy: If the weather forecast states that it will be sunny tomorrow without providing the probability, you feel lied to if it actually rains. But if you were informed that the chance of a sunny day is 80%, would you go out without an umbrella knowing there is also a 20% chance of rain? Context is everything and this needs to be discussed among clinicians and patients before agreeing on a certain threshold.

General practitioner's role

Thus, PCa testing involves a complex individually-based decision-making process. This process starts with the general practitioner (GP), where the GP should first inform the patient regarding the pros and cons of PCa testing. The decision-making involves accounting for the unexpected or non-preferable scenarios that may have serious psychological implications [46]. For example, if a man decides not to test for PCa but develops back pain caused by PCa metastases later on, he might be angry with himself or the doctor, although it remains uncertain whether he would have benefited from PCa screening. And vice versa, if a patient was screened for PCa and PCa was detected at an early phase; he may think screening was beneficial.

The individually based decision-making process for PCa is also complex due to unavailable information to provide a proper context. For example, information on quality of life in the screening setting is important, however this is often unavailable. Furthermore, metastatic findings having bearing on end-of-life experiences are underreported in screening trials. To develop ground rules for the shared decision-making process, we used PCa mortality as this was the main outcome in the ERSPC study and is available in public registries. With this information, we were able to use life expectancy as an outcome in order to provide insights when making a decision to refer to a urologist for a prostate biopsy or not. As most PCa grows slowly, estimating the survival benefit (years gained) from the early detection of asymptomatic clinically significant (cs)PCa should be considered in information-sharing in the context of competing risks from other life-threatening comorbidities. This makes the decision to test for PCa even more complex, as both the life expectancy after PCa treatment and other concurrent diseases have to accounted for. Chapter 7 describes an integrated approach weighing these aspects. The proposed tool can help primary care physicians in patient triage for timely and necessary referral for further assessment, and as such, can aid in reducing unnecessary testing, overdiagnosis, and subsequent overtreatment [47]. It reduces the possibility of unnecessary referrals of elderly men simply on the basis of having an elevated risk of having a csPCa, for example men with high PSA level. This corresponds with guidelines advising not to refer men aged >75 as they have a low life expectancy [48-50]. This tool was developed using the highest qualified and currently available data sources, but statistical assumptions were required to combine these different data sources; a possible limitation. Therefore, this tool needs to be validated more thoroughly in other settings before being implemented in daily practice. Only then is a full answer possible with respect to the extent to which this model can support triage in primary care practice regarding who deserves screening and diagnostic examination, thereby reducing unnecessary testing and overdiagnosis.

CONCLUSIVE SUMMARY

In this thesis, five research questions were answered related to two major topics. The first topic addressed the research results of prostate cancer screening and active surveillance. It explored the translation of the relative risk reduction of the screening into the real-life scenario of the absolute risk reduction by risk-stratification. The second topic evaluated the subject of PCa screening from a personalized medicine perspective. More specifically we asked: can we reduce unnecessary testing using multivariable prostate cancer prediction models, and can risk-stratified models be of value in clinical decision-making?

Classic evidence-based medicine: Outcomes of prostate cancer screening and active surveillance

Question 1 – *What are the risks of a diagnosis of a clinically significant PCa, metastatic disease and/or PCa death after a false negative screening test or biopsy result in a purely PSA-based screening setting compared to applying an additional test procedure or risk stratification tool? (Chapter 2)*

There is strong evidence that PSA screening by applying systematic prostate biopsies for test positivity criterion ≥3.0 ng/mL is beneficial in achieving a 20 % relative risk reduction of PCa death. A logical question is whether a larger benefit is possible by addressing the false negative rates for men with PSA <3.0 ng/mL and negative biopsy. We discovered that these false negative rates are extremely low and PSA (repeat) screening (including sextant biopsies) detects almost every PCa case. Therefore, additional tools like biomarkers and mpMRI are unlikely to detect more clinically relevant PCa. Nonetheless, these additional tools may be useful for risk-stratification and have added value in reducing the high rates of unnecessary biopsy (no cancer) and considerable overdiagnosis of low-risk PCa.

Question 2 – *What is the compliance over time when offering an AS protocol to men with low-risk PCa and how can risk stratification at the start of AS optimize adherence? (Chapter 3)*

AS is based on the concept that low-risk PCa is unlikely to harm or decrease life expectancy. Management of slow growing PCa with AS is a better choice than immediate active treatment with surgery or radiation, including complications and side effects. The strategy is substantiated by studies that show that men with low-risk prostate cancer who have been on AS for 10 - 15 years after diagnosis have extremely low rates of disease spread or death from prostate cancer. In addition to the definition of low-risk and/or indolent PCa, the eligibility and inclusion criteria for AS, the AS protocol itself, and adherence to the protocol is key determinant of progression of disease.

The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance Initiative (GAP3 initiative) provided results from 25 AS cohorts worldwide. Adherence to the annual schedule of repeat biopsy reduced with time: we estimated a pooled one-third nonadherence after 7 years. For making evidence-based AS selection guidelines, the importance of risk stratification with age, tumor stage, PSA level, and diagnosed biopsy score was stressed to optimize future AS protocols, mpMRI surveillance was also considered. Overall, constructing a more personalized risk-based approach to biopsy scheduling on AS was recommended.

Towards personalized medicine: How can multivariable prostate cancer prediction models reduce unnecessary testing and support clinical decision-making?

Question 3 – *How do prediction models work and how should the prediction outcomes be interpreted in terms of clinical utility? (Chapter 4)*

Classical evidence-based medicine methods analyze the benefits and harms of tests and treatments in terms of relative risk reduction and/or elevation. The concept of numbers needed to screen, biopsy, diagnose, treat, and harm is used as a better way of communicating and clinical decision-making. A number needed to screen of, say, 570 means that the death of 1 in 570 men will be prevented through PCa, and the remainder will not. Unfortunately, we do not yet know who the 'lucky' one will be. Prediction models address the unique situation for the individual man. These models are designed and include patient characteristics, test results, biomarkers and imaging to accurately predict the occurrence of defined endpoints. The clinical impact of newly developed risk prediction models is currently assessed with decision curves.
In the 'statistics in urology review' presented in Chapter 4, we made a plea for investigators with regards to reporting and correctly interpreting decision curve analysis. We present a statistical method to evaluate whether the model is useful in clinical decisions, and whether extended models with, for example, innovative biomarkers to predict high-grade PCa, will lead to better decisions.

Question 4 – *Can prediction model predicting biopsy outcome be improved by incorporating novel biomarkers and a more refined PCa pathological grading system, and hence decrease the number of unnecessary prostate biopsies and overdiagnosis of potentially indolent disease? (Chapters 5 and 6)*

Strong evidence that primary-based PSA testing works to reduce PCa mortality has been presented. However, side effects are unavoidable because a raised test result does not necessarily mean a PCa diagnosis, so further examination is needed. The PSA test has a high

false positive rate (the predictive value positive of prostate biopsy is low), indicating a situation of too many unnecessary biopsies.
Prediction models are adopted to refine diagnosis (differentiation of potentially lethal PCa from relatively indolent cancer) and reduce the number of unnecessary biopsies. The models incorporate patient characteristics, biomarkers and technology to optimize the balance between the likelihood of benefit and the risk of harm for individuals. We studied the prediction model underlying the No. 3 Rotterdam European Randomized Study of Screening for Prostate Cancer risk calculator for initial prostate biopsy and compared it with an updated risk calculator augmented with the contemporary Gleason grading and cribriform growth pathological biopsy classification. This Rotterdam calculator will lead to a 34% reduction in unnecessary biopsy, while only 2% of high-risk PCa will be undiagnosed.
We then searched for a novel biomarker to detect PCa to be included as a risk calculator able to deliver optimal net benefit outcome, the 4-kallikrein panel (4K-score). The prediction performance was studied with discrimination and calibration plots, and decision curve analysis was used to evaluate the reduction of unnecessary biopsy and indolent PCa. Compared to PSA testing, the Rotterdam Prostate Cancer Risk Calculator and the 4K-score equally reduce the number of biopsies by approximately two-thirds. Prostate cancer prediction can be slightly improved by combining the Rotterdam calculator with the 4K-score.

Question 5 – *To what extent can prediction models support triage at primary care practice regarding who receives screening and diagnostic examination, thereby reducing unnecessary testing and overdiagnosis? (Chapter 7)*

Current consensus-based guidelines advise not to refer men for biopsy if aged >75 as they have a low life expectancy. When making these decisions, it is relevant to know that 50% of Dutch 75-year-old men are expected to live more than 11.5 years, but also that approximately 25% of them will live for more than 15 years, while only 25% will live less than 6 years. We therefore recommend adding to clinical usefulness of the tool's information on risk of cancer and life expectancy.

A tool for shared-decision-making on referral for biopsy in the primary care setting was developed. Data and estimates were collected from the Dutch arm of the ERSPC trial, treatment trials, cancer registries, and national mortality statistics. A negative impact on life expectancy and treatment benefit was found with higher age and more comorbidity. The proposed multivariable and multidimensional prediction tool with information on life expectancy, risk of aggressive PCa, and potential benefit of prostate cancer treatment comorbidity needs further validation. We are confident that it can provide general practitioners and their patients with more accurate information regarding whether or not the patients should be referred for prostate biopsy.

FUTURE PERSPECTIVES

Developments in the early detection of PCa and prostate cancer care go fast. Classic evidence-based medicine applies to the average population and cannot simply be translated into person-level predictions, which is the objective of personalized medicine. How the future will unfold can only be presented in general terms. Table 2 provides summary points.

Table 2. Summary points

- Classic evidence-based medicine applies to the average population and cannot simply be translated into person-level predictions
- The current state of evidence on prostate cancer screening is still regarded insufficient to start large-scale screening programs
- Updates with novel biomarkers and imaging techniques could favor the discussion to start with screening programs and should be weighed with other significant improvements in terms of operating techniques, systemic treatments, localized radiology treatment as they also influence PCa mortality
- Active surveillance is a safe option for men with low-risk PCa, however the definition of low-risk PCa should be defined in absolute risks of clinically significant PCa
- The shift towards personalized medicine with prediction models provides more patient-specific intervention estimates which support individualized clinical decision-making, instead of using a relative risk of intervention.
- Decision curve analysis is introduced as a novel method for evaluating the clinical usefulness of prediction models to aid patients with their decisions, however, there is still a way to go.

Prostate cancer screening

The current state of affairs and body of evidence on prostate cancer screening is still regarded insufficient to start large-scale screening programs [51]. Still, prediction models that include biomarkers and imaging findings are promising and increase the net benefit of prostate cancer screening, mainly by reducing the associated harms. However, it has to be kept in mind that most evidence is still based on small population and patient cohorts, and often on retrospective studies. In recent years, many companies have developed novel biomarkers such as the 4Kscore (OPKH Health Inc), PHI (Beckman Coulter), SelectMDx (MDxHealth Inc), and ExoDX (Bio-Techne). In addition, there have been great improvements in histopathological features and imaging techniques like TRUS, PSMA-PET/CT and mpMRI techniques (Philips, Siemens and General Electrics). Before a proper screening program can be implemented, hard evidence on the predictability of these biomarkers, histopathological

features and imaging techniques needs to be gathered based on new, large screening trials powered on intermediate endpoints or cohort studies opting for predicting modeling and personalized medicine. Moreover, these predictors should be compared with each other to improve the total clinical utility. A new Finnish screening study is currently under way, conducted with 4Kscore and mpMRI [52], but before the results can be interpreted more participants and longer follow-up are required [53].

It should be realized that our understanding of the benefits of prostate screening comes from 25-year-old data, and updates are strongly needed, as significant improvements in terms of operating techniques, systemic treatments, localized radiology treatment also influence PCa mortality. As times moves on, even the data resulting from a new screening trial will become outdated, thus instead of continually initiating a new trial, better registration of large cohorts is a good alternative.

Cost-effectiveness considerations are also important, not only when making decisions regarding launching a population-wide screening program, but also in individual testing. Studies comprising cost-effectiveness analysis are required when improvements in clinical decision-making are expected.

Active surveillance

AS is a safe option for men with low-risk PCa, however the definition of low-risk PCa should be defined in absolute risks of clinically significant PCa, indolent cancer or benign hyperplasia instead in relative risk reductions to provide appropriate care for each individual. In the near future, clinical and histopathological features, biomarkers, and imaging techniques should be used in a complementary manner in multivariable prediction models. These models may achieve optimal risk stratification and maximize the effects, e.g. the avoidance of detection low-risk PCa.

Prediction models

We now live in an era where vast amounts of data can be stored and quickly processed, with new advances in developing and validating accurate prediction models. What was first a simple linear correlation can now be a higher order transformation with restricted cubic spline with multiple knots, or even more flexible exploitation of correlation with machine learning methods and artificial intelligence. However, we should be aware of the problems related to overfitting which occurs when a prediction model is too complex to be developed in a specific sample of limited size. Moreover, transportability may be limited when a model is implemented a clinical setting other than the setting for which it was developed [10]. Overfitting can be prevented by using large numbers in the development of the prediction model, but also with sensible modeling [54]. Sensible modeling means applying external knowledge and using a model only for a specific calculation. Also, as overfitting automatically occurs shrinking the model should be applied. This is done by estimating the overfitting factor with cross-validation or bootstrapping techniques. Finally, external validation is used to see

the heterogeneity between different settings, if the heterogeneity is low the model is well transportable. Instead of statistically measuring how well a prediction model works, the clinical outcome can be simulated before implementation, the first step in this process is possible by using the Net Benefit approach [27]. The shift towards personalized medicine provides more patient-specific intervention estimates which support individualized clinical decision-making, instead of using a relative risk of intervention [55]. Therefore, healthcare with predictions models will benefit patients and provide an overall higher clinical utility.

Aiding patients with their screening decisions

In modern medical practice, decisions and interventions can benefit from the results of comparative group research. Using group comparison, it is possible to quantify the relative benefit for the individual. For example, the number needed to screen indicates how many people need to comply with screening in order to avoid one PCa cancer death. As a doctor aiding his or her patients with the decision to screen or not to screen, more information is needed. Decision curve analysis was introduced as a novel method for evaluating the clinical usefulness of prediction models, and adopted gradually by the urological community. However, there is still a way to go. The initial authors developed the tutorial because investigators indicated that decision curve analysis is difficult to understand, most probably as “the two axes of the decision plot —threshold probability and net benefit— are concepts that are novel to many” [56]. They argued that “many of the difficulties in interpreting decision curves can be solved by relabeling the y-axis as “benefit” and the x-axis as “preference”, i.e. a new x-axis ranging from “I’m worried about disease” towards “I’m worried about biopsy”. Still uncertainty about the future is considered to be the most common source of stress in humans apart from traumatic stress and can be difficult to put in the perspective of risk prediction and risk management for most people. Time will tell how instruments for shared-decision-making will evolve.

REFERENCES

1. Hugosson J, Roobol MJ, Mansson M, Tammela TLJ, Zappa M, Nelen V, et al. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. European urology. 2019.
2. Schröder FH, Hugosson J, Roobol MJ, Tammela TLJ, Zappa M, Nelen V, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. The Lancet. 2014;384:2027-35.
3. Roobol MJ, Kranse R, Bangma CH, van Leenders AG, Blijenberg BG, van Schaik RH, et al. Screening for prostate cancer: results of the Rotterdam section of the European randomized study of screening for prostate cancer. European urology. 2013;64:530-9.
4. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, et al. Systematic review of complications of prostate biopsy. European urology. 2013;64:876-92.
5. Verbeek JFM, Roobol MJ, group ERs. What is an acceptable false negative rate in the detection of prostate cancer? Transl Androl Urol. 2018;7:54-60.
6. Moussa AS, Meshref A, Schoenfield L, Masoud A, Abdel-Rahman S, Li J, et al. Importance of additional "extreme" anterior apical needle biopsies in the initial detection of prostate cancer. Urology. 2010;75:1034-9.
7. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. N Engl J Med. 2018;378:1767-77.
8. Drost FH, Osses DF, Nieboer D, Steyerberg EW, Bangma CH, Roobol MJ, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database Syst Rev. 2019;4:Cd012663.
9. Drost FH, Osses D, Nieboer D, Bangma CH, Steyerberg EW, Roobol MJ, et al. Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. European urology. 2019.
10. Verbeek JFM, Nieboer D, Steyerberg EW, Roobol MJ. Assessing a Patient's Individual Risk of Biopsy-detectable Prostate Cancer: Be Aware of Case Mix Heterogeneity and A Priori Likelihood. European urology oncology. 2019.
11. Randazzo M, Beatrice J, Huber A, Grobholz R, Manka L, Chun FK, et al. Is further screening of men with baseline PSA < 1 ng ml(-1) worthwhile? The discussion continues-Results of the Swiss ERSPC (Aarau). Int J Cancer. 2015;137:553-9.
12. Ploussard G, Nicolaiew N, Marchand C, Terry S, Vacherot F, Vordos D, et al. Prospective evaluation of an extended 21-core biopsy scheme as initial prostate cancer diagnostic strategy. European urology. 2014;65:154-61.
13. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality, Surveillance Research Program. April 2017.
14. Bokhorst LP, Bangma CH, van Leenders GJ, Lous JJ, Moss SM, Schröder FH, et al. Prostate-specific antigen-based prostate cancer screening: reduction of prostate cancer mortality after correction for nonattendance and contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. European urology. 2014;65:329-36.
15. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. Cochrane Database Syst Rev. 2013:CD004720.
16. Tosoian JJ, Carter HB, Lepor A, Loeb S. Active surveillance for prostate cancer: current evidence and contemporary state of practice. Nature reviews Urology. 2016;13:205-15.

17. Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med. 2016.

18. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. European urology. 2013;63:597-603.

19. Nyame YA, Almassi N, Haywood SC, Greene DJ, Ganesan V, Dai C, et al. Intermediate-Term Outcomes for Men with Very Low/Low and Intermediate/High Risk Prostate Cancer Managed by Active Surveillance. The Journal of urology. 2017;198:591-9.

20. Musunuru HB, Yamamoto T, Klotz L, Ghanem G, Mamedov A, Sethukavalan P, et al. Active Surveillance for Intermediate Risk Prostate Cancer: Survival Outcomes in the Sunnybrook Experience. The Journal of urology. 2016;196:1651-8.

21. JFM Verbeek, MJ Roobol, EW Steyerberg (2018). Risk-based selection for active surveillance In L Klotz (Ed.), Active surveillance for localized prostate cancer/2nd (pp. 53-64). Springer International Publishing

22. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. BJU Int. 2017;119:560-6.

23. Venderbos LD, van den Bergh RC, Roobol MJ, Schröder FH, Essink-Bot ML, Bangma CH, et al. A longitudinal study on the impact of active surveillance for prostate cancer on anxiety and distress levels. Psycho-oncology. 2015;24:348-54.

24. Van Calster B, Van Belle V, Vergouwe Y, Steyerberg EW. Discrimination ability of prediction models for ordinal outcomes: relationships between existing measures and a new measure. Biom J. 2012;54:674-85.

25. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. Med Decis Making. 2013;33:490-501.

26. Verbeek AJ, Verbeek JF, van Dijck JA, Verbeek AL. [Predicting chance of disease: calculation using prediction rules]Ziektekansen voorspellen: rekenen met predictieregels. Ned Tijdschr Geneeskd. 2014;158:A7041.

27. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-74.

28. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj. 2015;350:g7594.

29. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. European urology. 2018.

30. Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making. 2015;35:162-9.

31. Loeb S, Bjurlin MA, Nicholson J, Tammela TL, Penson DF, Carter HB, et al. Overdiagnosis and overtreatment of prostate cancer. European urology. 2014;65:1046-55.

32. Moynihan R, Nickel B, Hersch J, Beller E, Doust J, Compton S, et al. Public Opinions about Overdiagnosis: A National Community Survey. PLoS One. 2015;10:e0125165.

33. Howard K, Salkeld GP, Patel MI, Mann GJ, Pignone MP. Men's preferences and trade-offs for prostate cancer screening: a discrete choice experiment. Health Expect. 2015;18:3123-35.

34. Roobol MJ, Schröder FH, Hugosson J, Jones JS, Kattan MW, Klein EA, et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC)

risk calculators: results from the prostate biopsy collaborative group. World J Urol. 2012;30:149-55.

35. Roobol MJ, Verbeek JFM, van der Kwast T, Kummerlin IP, Kweldam CF, van Leenders G. Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculator for Initial Prostate Biopsy by Incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth. European urology. 2017;72:45-51.
36. Pruthi DK, Ankerst DP, Liss MA. Novel Definitions of Low-risk and High-risk Prostate Cancer: Implications for the European Randomized Study of Screening for Prostate Cancer Risk Assessment Tool. European urology. 2017.
37. Kweldam. CF, Kümmerlin. IP, Nieboer. D, Verhoef EI, Steyerberg. EW, Incrocci. L, et al. Prostate cancer outcomes of men with biopsy Gleason score 6 and 7 without cribriform or intraductal carcinoma. European Journal of Cancer. 2016.
38. Kweldam CF, Kummerlin IP, Nieboer D, Verhoef EI, Steyerberg EW, van der Kwast TH, et al. Disease-specific survival of patients with invasive cribriform and intraductal prostate cancer at diagnostic biopsy. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2016;29:630-6.
39. Klotz L, Vesprini D, Sethukavalan P, Jethava V, Zhang L, Jain S, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol. 2015;33:272-7.
40. Vickers AJ, Cronin AM, Aus G, Pihl CG, Becker C, Pettersson K, et al. 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 Med. 2008;6:19.
41. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21:128-38.
42. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol. 2018;98:133-43.
43. Van Calster B, Steyerberg EW, Harrell FH. Risk Prediction for Individuals. Jama. 2015;314:1875.
44. Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. New York: Springer Sciences & Business Media, 2019.
45. Kattan MW, Eastham JA, Wheeler TM, Maru N, Scardino PT, Erbersdobler A, et al. Counseling men with prostate cancer: a nomogram for predicting the presence of small, moderately differentiated, confined tumors. The Journal of urology. 2003;170:1792-7.
46. Reyna VF, Nelson WL, Han PK, Pignone MP. Decision making and cancer. The American psychologist. 2015;70:105-18.
47. Verbeek JFM, Nieboer D, Parker C, Kattan MW, Steyerberg EW, Roobol MJ. A Tool for Shared Decision Making on Referral for Prostate Biopsy in the Primary Care Setting: Integrating Risks of Cancer with Life Expectancy. Journal of personalized medicine. 2019;9.
48. Cooperberg MR. Implications of the new AUA guidelines on prostate cancer detection in the U.S. Curr Urol Rep. 2014;15:420.
49. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. European urology. 2017;71:618-29.
50. NVU richtlijn https://richtlijnendatabase.nl/richtlijn/prostaatcarcinoom/algemeen.html.

51. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69:184-210.

52. Auvinen A, Rannikko A, Taari K, Kujala P, Mirtti T, Kenttamies A, et al. A randomized trial of early detection of clinically significant prostate cancer (ProScreen): study design and rationale. Eur J Epidemiol. 2017;32:521-7.

53. Rannikko A, Kilpeläinen T, Matikainen M, Kenttämies A, Petas A, Santti H, et al. 1135 - A population-based randomized trial of early detection of clinically significant prostate cancer (ProScreen): Pilot phase results. European Urology Supplements. 2019;18:e1535-e6.

54. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-7.

55. Kent DM, Steyerberg E, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. Bmj. 2018;363:k4245.

56. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res. 2019;3:18.

SUMMARY

The western world has a high burden of disease when it comes to prostate cancer. In 2018 in the Netherlands, 12,646 men were diagnosed with and 2894 died from prostate cancer [1]. The last three months of the prostate cancer disease process are tough, first and foremost for the terminally ill patients, but also for their family, informal caregivers and medical-professional caregivers. The desire for both early detection of prostate cancer and better treatment by testing for the prostate-specific antigen (PSA) in the blood to enhance the cure rate is therefore logical. Currently, the Netherlands does not support a prostate cancer population screening program comparable to the 2-yearly breast cancer screening for women aged 50-75.

Three-quarters of all prostate carcinomas are discovered with the PSA test. Males usually request PSA testing when they have symptoms related to the lower urinary tract or miction complaints, or if they suspect cancer [2]. In the Netherlands, the PSA test is included in clinical guidelines for urologists as well as in GPs' NHG standard [3]. In the United States, the American Cancer Society states the following when it comes to prostate cancer screening: "… Men who have at least a 10-year life expectancy should have an opportunity to make an informed/shared decision with their health care provider about whether to be screened for prostate cancer with serum prostate-specific antigen (PSA), with or without digital rectal examination (DRE), after receiving information about the benefits, risks, and uncertainties associated with prostate cancer screening and therapy" [4].

The benefits, harms and uncertainties around early diagnosis and timely treatment cannot be easily captured purely in numbers, specifically as it can vary per individual case. To improve their grip on this, the American Cancer Society article continues, stating: "… methods have become available that merge this information to achieve an estimate of a man's overall risk of prostate cancer and, more specifically, his risk of high-grade prostate cancer. These risk-stratification/decision-making algorithms are intended to increase the benefit of testing and reduce the harms associated with biopsy and treatment of low-risk prostate cancer."
This statement refers to the so-called risk calculators that are gradually finding their way in clinical practice; see for example www.prostaatwijzer.nl/medical-risk-calculators. Use of these prostate risk calculators should lead to fewer biopsies. Currently, the accuracy of the predictions and the improvement of the calculators is the subject of much screening research and patient populations, these risk calculators integrated findings of rectal toucher, PSA parameters, other biomarkers, ultrasound and/or MRI.
However, active treatment is not always required after prostate cancer diagnosis. In many cases, watchful waiting is recommended, with additional arguments for active surveillance with the additional help of risk calculators, especially for low-grade localized prostate cancer.

In this thesis, I describe the results of my research into the clinical utility of existing risk calculators and their updates. I also discuss the effect of the underlying patient risk of low to high ranging prostate cancer, and the life expectancy of the case in question, focussing on middle-aged and elderly males.

In **Chapter 1**, I first discuss the functions of the prostate gland, its pathophysiology, and the development of prostate cancer. The Netherlands faces a high prostate cancer disease burden, with 12,646 men diagnosed in 2018 and a high mortality with 2862 death [1]. At the individual level, the incidence rate means that 1 in 8-11 Dutch men will be affected by prostate cancer during their lifetime. Incidence has stabilized in the last two decades and mortality rates have decreased. Figure 1, shows that age-specific incidence rates have fallen in the last two decades and that the peak age occurs slightly earlier (70th year) in 2018. Also, the mortality rates have clearly decreased for all age categories over the period 1998-2017.

The history of PSA blood measurements shows the evolution of PSA testing; from setting a diagnosis and clinical follow-up examinations to early detection, which has led to the start of numerous studies into the effects of mass screening. The large-scale and longstanding European Randomized Study of Screening for Prostate Cancer (ERSPC) project shows a reduction of prostate cancer mortality, but also demonstrates the occurrence of overdiagnosis of indolent prostate cancer as a result of PSA screening. Active surveillance (AS) counteracts the over-treatment of these harmless forms of prostate cancer. AS should be the first treatment strategy in the intensive follow-up of men with low-risk prostate cancer and adequate life expectancy, until disease progression is established and treatment with curative intent is initiated. Prediction models have been introduced that combine patient and test results to estimate the probability of detecting low-risk prostate cancer with the aim of reducing overdiagnosis and overtreatment. With this thesis, I set out to contribute to the development, evaluation and implementation of prediction models in the early detection of prostate cancer.

Currently, the Netherlands has not introduced a prostate cancer screening program as it is said that the benefits (reduction of prostate cancer mortality) do not outweigh the disadvantages (among others, overdiagnosis of indolent prostate cancer). If a man visits a GP for consultation, whether or not he needs testing for prostate cancer, a well-informed assessment should first be made in relation to the benefits and harms of PSA testing. Testing with PSA and performing prostate biopsies is the most common way to detect prostate cancer. However, it remains unclear what kind of follow-up is required when negative PSA test results (PSA $<$3.0 ng/ml) or negative biopsies (prostate biopsies with no prostate cancer) are found. In **Chapter 2**, I describe the possible consequences of a false negative result based on empirical data from the ERSPC study. In this randomized trial, a few false negative findings

were reported with the PSA test and biopsies. Improved risk stratification with prediction models may possibly result in the detection of more high-risk prostate cancers while not requiring an increase in biopsies.

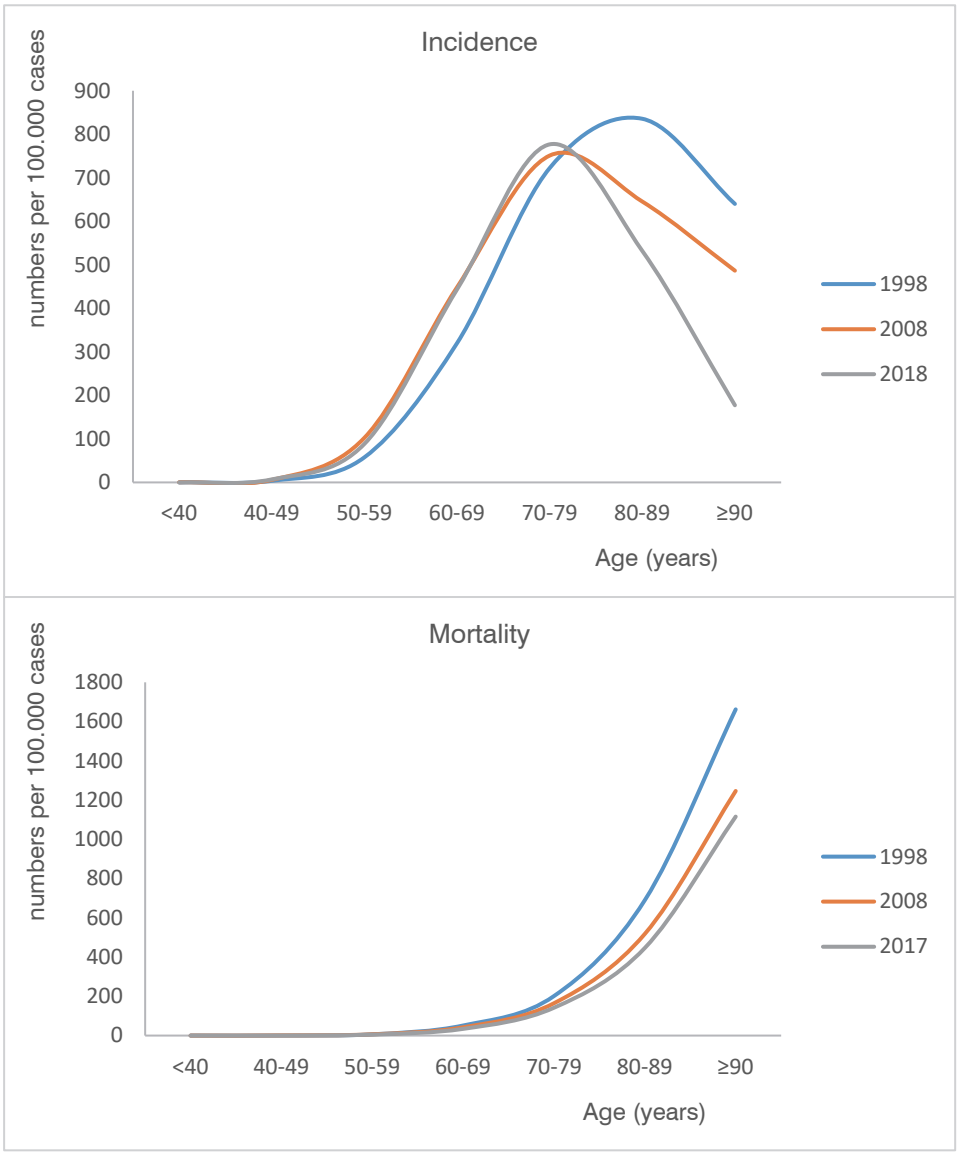


Figure 1. Age-specific incidence and mortality for three calendar years [1]

Since the metamorphosis of AS from an ad hoc strategy into a routine and formal practice, AS has been established as a key management strategy for localized low-risk prostate cancer. However, AS enrolment criteria and follow-up schedules for low-risk prostate cancer can vary between institutions and health care settings, leaving uncertainty about adherence to these protocols. In **Chapter 3**, adherence to institution-specific AS inclusion criteria and follow-up schedules within the Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) initiative were studied. Inclusion criteria were generally applied well, but adherence to repeat biopsy decreased with time. The patient's compliance needs to be considered to optimize AS protocols in the future, and it is important to apply risk stratification right from the start of AS.

The predictive performance of a prediction model is originally evaluated by statistical measures including discrimination and calibration. However, these standards do not provide sufficient insights into the added value of using a (new) prediction model compared with the standard health care strategy. In **Chapter 4**, I introduce the concept of Net Benefit (NB) incorporated in Decision Curve Analysis (DCA). I explain how NB and DCA are able to analyze the added value of using prediction models to aid the process of clinical decision making for PSA testing, biopsy taking and AS.

In **Chapter 5**, I dig deeper into the use of NB and DCA. The survival rate of patients with International Society of Urological Pathology (ISUP) grade 2 prostate cancer without invasive cribriform (CR) or intraductal carcinoma (IDC) growth was found to be similar to that of men with ISUP grade 1 (innocent form of) prostate cancer. This histological finding was added to the existing ERSPC risk calculator #3, with showing a further reduction of unnecessary prostate biopsies while still detecting aggressive (high-risk) forms of prostate cancer.

In **Chapter 6**, I describe the effects of adding the 4Kscore to the updated ERSPC risk calculator including the latest ISUP grading and the secondary growth patterns CR and IDC. In this study, both the 4K score and the ERSPC-RC showed equal clinical utility, while the addition of the 4Kscore slightly improved clinical utility. However, this improvement of reducing unnecessary biopsies should be weighed against the costs of the 4Kscore.
In **Chapter 7**, I discuss the complex individual decision-making process for a GP with his patient regarding the possible referral of a patient with suspected prostate cancer to the urologist for prostate biopsies. In this decision-making process, the expected survival benefit from early detection of prostate cancer should be higher than the risk of death from other causes (comorbidity). By developing a prediction model based on the outcomes of a number of large international studies from different health care settings, and taking life expectancy into account, I make specific recommendations to assist the decision-making process regarding diagnostic testing and screening for prostate cancer.

REFERENCES

1. Nederlandse kankerregistratie: cijfers over kanker. Accessed through: www.cijfersoverkanker.nl on Feb 2019.
2. Blanke M.H, Bangma C.H. Wel prostaatkanker, maar niet actief behandelen. Ned Tijdschr Geneeskd. 2019;163:D3698.
3. NVU richtlijn https://richtlijnendatabase.nl/richtlijn/prostaatcarcinoom/algemeen.html.
4. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69:184-210.

SUMMARY IN DUTCH

Net als in andere westerse landen kent prostaatkanker in Nederland een grote ziektelast. In 2018 werd 12.646 keer de diagnose gesteld en was het aantal overlijdens door prostaatkanker 2894 [1]. Voor de terminale patiënt zijn de laatste maanden van het ziekteproces zwaar, ook voor zijn familie, mantelzorgers en medisch-professionele hulpverleners. Het is dan ook logisch dat men naar verbetering van therapiemodaliteiten kijkt en naar mogelijkheden van vroege ontdekking van prostaatkanker door prostaat-specifieke antigeen (PSA) in het bloed te bepalen om langs deze weg de genezingskans te verhogen. In Nederland is geen bevolkingsonderzoek naar prostaatkanker zoals dat wel het geval is voor borstkanker met mammografische screening elke 2 jaar voor vrouwen van 50–75 jaar oud.

Driekwart van de prostaatcarcinomen wordt gevonden via de PSA-test. Het is meestal de man zelf die om bepaling van de PSA-waarde vraagt bij symptomen van de lagere urineweg, mictieklachten of anderszins bij verdenking op kanker [2]. In Nederland is het gebruik van de PSA-test vervat in de klinische richtlijn van de urologen en opgenomen in de NHG-standaard [3]. In de V.S. heeft de American Cancer Society een specifieke richtlijn voor prostaatkanker-screening en meldt het volgende (vertaald uit het Engels): "… Alle mannen met een levensverwachting van minstens 10 jaar dienen in de gelegenheid te worden gesteld met een arts te overleggen of het zinnig is zich met de PSA-test te laten screenen op prostaatkanker al dan niet met rectaal toucher, waarbij hij niet alleen informatie krijgt over het gunstige effect van screening, maar ook over de risico's en de onzekerheden van prostaatkankerscreening en therapie" [4].

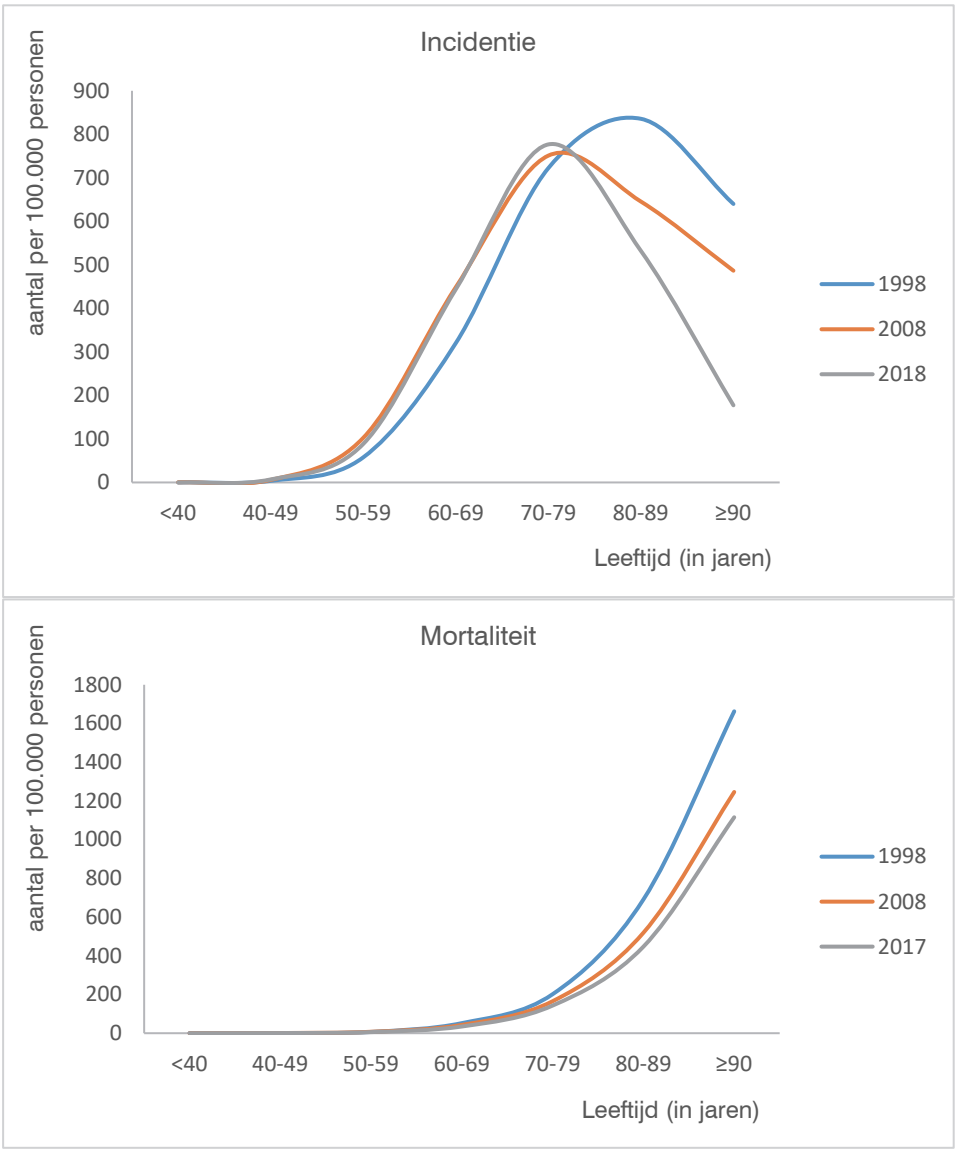
De voordelen, nadelen en onzekerheden rond vroege diagnostiek en tijdige behandeling zijn niet in een paar getallen te vatten en kunnen ook voor de individuele man anders uitpakken. Om hierop greep te krijgen, zo vervolgt het artikel van de American Cancer Society, zijn er methoden beschikbaar gekomen (vertaald uit het Engels): "… die alle deelinformatie bij elkaar voegt om een goede inschatting te krijgen van het risico dat de desbetreffende man loopt om prostaatkanker te krijgen en, meer specifiek, het risico op een hooggradig carcinoom. Deze risico-algoritmes zijn bedoeld om op individueel niveau het voordeel van een PSA-test te verhogen en de nadelen van het testen zoals het nemen van biopten en het behandelen van laagrisico prostaatkanker te verminderen."
Het gaat hierbij om de zo genoemde risico-calculators die op basis van de resultaten van velerlei medisch-wetenschappelijk onderzoek zijn ontwikkeld en in toenemende mate in de praktijk gebruikt worden; zie www.prostaatwijzer.nl/medical-risk-calculators. Toepassing van de prostaatwijzer leidt tot minder vaak biopteren, omdat bijvoorbeeld de noodzaak daartoe bij individuen getalsmatig lager wordt ingeschat. De nauwkeurigheid van de voorspellingen

en het verbeteren van de calculators is onderwerp van veel research bij gescreende populaties en patiëntpopulaties met de bevindingen van rectaal toucher, PSA-parameters, andere biomarkers, echografie en/of MRI geïntegreerd. Dit geldt ook voor de prognostische kenmerken van het carcinoom als histopathologie en maligniteitsgraad. Na het stellen van de diagnose prostaatkanker hoeft niet altijd een actieve behandeling te volgen [2]. In veel gevallen is actief volgen of waakzaam afwachten zeer zeker te verdedigen met name voor laaggradig lokaal gelokaliseerde prostaatkanker. In deze thesis wordt het voorspellend vermogen van bestaande risico calculators en van de updates hiervan bestudeerd. Tevens wordt de impact van het onderliggende risico op prostaatkanker onderzocht, dat bij individuen kan variëren van erg laag tot hoog, en de levensverwachting van de man of patiënt in kwestie, doorgaans van middelbare of gevorderde leeftijd.

In **Hoofdstuk 1** worden de basisbegrippen geïntroduceerd die van belang zijn om de omvang van de effecten en bijeffecten te schatten wanneer men prostaatkanker vroeg wil ontdekken. Eerst komt kort de functie van de normale prostaat aan bod, gevolgd door de pathofysiologie en de ontstaanswijze van prostaatkanker. De beschrijvende epidemiologie van prostaatkanker laat vervolgens zien (zie onderstaande figuur) dat de leeftijdsspecifieke incidentiecijfers de laatste twee decennia lager zijn geworden en dat de piekleeftijd iets eerder optreedt, nl. tegen het 70e levensjaar in 2018 [1]. Met deze cijfers is te berekenen dat tegenwoordig ongeveer 1 op de 8 tot 11 mannen prostaatkanker krijgt. Het tweede deel van de figuur toont de sterftecijfers voor prostaatkanker in Nederland. De mortaliteitscijfers zijn over de periode 1998-2017 voor alle leeftijdscategorieën duidelijk afgenomen.

Vervolgens komt de geschiedenis van de meting van PSA in het bloed aan de orde, hetgeen heeft geleid tot de start van verschillende studies naar het effect van screening op prostaatkanker. Het grootschalige European Randomized Study of Screening for Prostate Cancer (ERSPC) is een langdurig project met een gerandomiseerde studie-opzet waardoor in principe weinig bias in de effectschatting van de screening optreedt. De ERSPC toonde aan, ook met vervolgstudies, dat de mortaliteit van prostaatkanker afneemt, maar tevens dat er overdiagnostiek plaatsvindt. Dit is de detectie van onschuldige vormen van prostaatkanker door te screenen met PSA, tumoren die buiten screening of bevolkingsonderzoek om niet aan het daglicht zouden treden. Om te voorkomen dat door behandeling van deze onschuldige vormen van prostaatkanker overbehandeling optreedt, is het niet-behandelen maar actief volgen – active surveillance (AS) – een optie voor mannen bij wie via een PSA-test prostaatkanker is vastgesteld, maar waarvan het erg onwaarschijnlijk is dat dit carcinoom ooit klachten zou geven. Met AS door middel van regelmatige controles worden deze mannen met laagrisico prostaatkanker en anderszins adequate levensverwachting stringent in de gaten gehouden. Mocht de tumor van aard of maligniteitsgraad verergeren of wanneer het risico op latere uitzaaiing toch hoger blijkt te zijn, dan kan de uroloog alsnog een in opzet

curatieve behandeling starten. Voor de tijdige inschatting van deze risico's zijn predictiemodellen ontworpen en geïntroduceerd, waarmee patiëntbevindingen en testresultaten worden gecombineerd waardoor op geleide hiervan een reductie van de aanvankelijke overdiagnostiek en overbehandeling kan worden gerealiseerd. Het doel van de thesis is om bij te dragen aan de verdere ontwikkeling, evaluatie en toepassing van predictiemodellen bij de vroege detectie van hoog risico prostaatkanker.



Figuur 1. Leeftijdsspecifieke incidentie en mortaliteit voor drie kalenderjaren. (Bron: via [1])

Momenteel is er in Nederland geen bevolkingsonderzoek naar prostaatkanker aangezien de voordelen (reductie van prostaatkankersterfte) worden ingeschat niet op te wegen tegen de nadelen zoals de overdiagnose van onschuldige vormen van prostaatkanker. Als een man nu zijn huisarts consulteert met de vraag of hij zich moet laten testen op prostaatkanker, zal bijgevolg een goed geïnformeerde afweging moeten worden gemaakt over de consequenties van de PSA-test. Momenteel is het testen met PSA en het biopteren van de prostaat de voornaamste manier om prostaatkanker te detecteren. Er is echter nog onduidelijkheid over de diagnostische, prognostische en therapeutische consequenties wanneer de PSA-test volgens het huidige criterium negatief is, dat willen zeggen bij PSA<3.0 ng/ml, of in geval van negatieve biopten (prostaatbiopten waarin geen prostaatkanker wordt aangetoond).

In **Hoofdstuk 2** worden de potentiële consequenties van deze fout-negatieve testuitslagen beschreven met empirisch datamateriaal van de ERSPC-studie. Uit deze studie kwam naar voren dat er weliswaar maar weinig fout-negatieve bevindingen zijn bij de PSA-test en met het nemen van biopten, maar dat de aantallen daarentegen niet te verwaarlozen zijn. Een goede risicostratificatie met predictiemodellen maakt het mogelijk om het aantal gemiste hoog-risico prostaatkankers verder te verminderen.

Met actief volgen van patiënten als een ad hoc toepassing naar een routinematige en protocollaire handelwijze, heeft AS zich langzamerhand ontwikkeld tot de eerste keus aanpak van het gelokaliseerde laagrisico prostaatkanker. Er is echter nog geen overeenstemming over de inclusiecriteria en het precieze follow-up schema voor mannen die AS krijgen. Om dit verder te adstrueren is in **Hoofdstuk 3** de naleving van ziekenhuis-specifieke AS-inclusiecriteria en follow-up schema's bepaald binnen het Global Action Plan van de Movember Foundation Prostate Cancer Active Surveillance (GAP3) initiatief. De conclusie is dat de inclusiecriteria goed worden gevolgd, maar dat de therapietrouw na herhaalde biopsies allengs minder wordt. Bij het optimaliseren van AS moet de therapietrouw van de patiënt worden meegenomen en is het belangrijk om risicostratificatie direct bij aanvang van AS toe te passen.

Het voorspellend vermogen van een predictiemodel wordt van oudsher weergegeven met statistische maatstaven als discriminatie en kalibratie. Deze maatstaven geven echter onvoldoende inzicht in wat nu de toegevoegde waarde is van het gebruik van een (nieuw) predictiemodel ten opzichte van de standaard strategie. In **Hoofdstuk 4** wordt daartoe het begrip Net Benefit (NB) uitgelegd dat een centrale plaats heeft in de Decision Curve Analysis (DCA) om zo toegevoegde waarde van een predictiemodel in de besluitvorming te kunnen analyseren. DCA is met diagrammen te visualiseren en kan tevens ondersteuning bieden bij de communicatie met de patiënt.

Hoofdstuk 5 gaat in op de overlevingskans van mannen met prostaatkanker met volgens de International Society of Urological Pathology (ISUP) graad 2 kenmerken en zonder cribriforme (CR) invasieve groei of intraductale carcinoom (IDC) groei. Patiënten met deze vorm van prostaatkanker hebben dezelfde overleving als mannen met ISUP-graad 1 prostaatkanker (een onschuldige vorm). In dit hoofdstuk wordt gevonden dat het toevoegen van deze histologische bevinding aan de bestaande ERSPC-risk calculator #3 het aantal te verrichten onnodige prostaatbiopten verder kan verminderen, terwijl de agressieve vormen van prostaatkanker nog steeds gedetecteerd worden.

In **Hoofdstuk 6** wordt geanalyseerd of het toevoegen van de 4Kscore predictor aan de reeks van ge-update ERSPC-riskcalculators voor een nog betere klinische utiliteit kan zorgen samen met de toevoeging van de vernieuwde ISUP-gradering inclusief secundaire groeipatronen (CR en IDC). In deze studie wordt gevonden dat zowel de 4Kscore als de ERSPC-RC gelijkwaardig zijn qua voorspellend vermogen en dat met toevoeging van de 4Kscore aan de ERSPC-RC een minimale verbetering te behalen valt. Deze verbetering in het verminderen van onnodige biopten dient echter nog wel afgewogen te worden met de kosten van de bepaling van de 4Kscore.

Hoofdstuk 7 behandelt de complexe individuele besluitvorming die een huisarts samen met een patiënt moet maken met betrekking tot al dan niet verwijzen naar de uroloog voor prostaatbiopten bij een mogelijke verdenking op prostaatkanker. In deze afweging moet de verwachte levenswinst van het vroeg detecteren van prostaatkanker groter zijn dan de kansen op overlijden door andere oorzaken, bijvoorbeeld vanwege co morbiditeit. Dit is een moeilijke inschatting voor huisarts en patiënt. Door de ontwikkeling van een predictiemodel op basis van verschillende grote internationale studies, met diverse populaties en levensverwachtingen, is getracht concrete adviezen te formuleren als hulpmiddel bij deze gezamenlijke besluitvorming.

REFERENCES

1. Nederlandse kankerregistratie: cijfers over kanker. Accessed through: www.cijfersoverkanker.nl on Feb 2019.
2. Blanke M.H, Bangma C.H. Wel prostaatkanker, maar niet actief behandelen. Ned Tijdschr Geneeskd. 2019;163:D3698.
3. NVU richtlijn https://richtlijnendatabase.nl/richtlijn/prostaatcarcinoom/algemeen.html.
4. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2019: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2019;69:184-210.

APPENDICES

ABOUT THE AUTHOR

Johannes (Jan) Franciscus Marcus Verbeek was born in Nijmegen on 11 February 1989. He completed his secondary education at the Nijmeegse Scholengemeenschap Groenewoud in 2007. He studied Biomedical Sciences at Radboud University Nijmegen, starting in 2007 and graduating with a Bachelor's degree in 2010. The close relationship between Biomedical Sciences and medicine stimulated him to study further; in 2010 he therefore decided to study medicine at Radboud University Nijmegen. While studying medicine he registered as Clinical Epidemiologist through the Netherlands Epidemiology Society. After being awarded his medical degree in 2015, he worked as resident at the department of Orthopedics at the Sint Maartenskliniek in Nijmegen. In 2016, he started his PhD project at the department of Urology at the Erasmus University Medical Center under the supervision of Prof. Dr. M.J. Roobol and Prof. Dr. E.W. Steyerberg. In 2019, he started working as resident at the Emergency department at Viecuri in Venlo. That's where his fascination with Radiology started. Since then, Jan has been working as ANIOS Radiology at the Radboud University Medical Center Nijmegen. In the future, he hopes to start as a radiology resident.

LIST OF PUBLICATIONS

List of publications by Jan Verbeek, spring 2020 (sorted by date of publication)
See also: https://www.researchgate.net/profile/Jan_Verbeek3

1. Chen R*, **Verbeek JFM***, Yang Y, Song Z, Sun Y, Roobol MJ. *Comparing the prediction of prostate biopsy outcome using the Chinese prostate cancer consortium (CPCC) risk calculator and the Asian adapted Rotterdam European randomized study of screening for prostate cancer (ERSPC) risk calculator in Chinese and European men.* World J Urol. 2020 Apr 11.

2. Osses DF, Drost FH, **Verbeek JFM**, Luiting HB, van Leenders GJLH, Bangma CH, Krestin GP, Roobol MJ, Schoots IG. *Prostate cancer upgrading with serial prostate magnetic resonance imaging and repeat biopsy in men on active surveillance: are confirmatory biopsies still necessary?* BJU Int. 2020 Mar 31.

3. Pullen L, Radtke JP, Wiesenfarth M, Roobol MJ, **Verbeek JFM**, Wetter A, et al. *External validation of novel magnetic resonance imaging-based models for prostate cancer prediction.* BJU Int. 2020;125:407-16.

4. Kalapara AA*, **Verbeek JFM***, Nieboer D, Fahey M, Gnanapragasam V, Van Hemelrijck M, et al. *Adherence to active surveillance protocols for low-risk prostate cancer: Results of the Movember foundation's global action plan prostate cancer active surveillance initiative.* Eur Urol Oncol. 2020;3:80-91.

5. **Verbeek JFM**, Roobol MJ. Reply by authors: *Reducing unnecessary biopsies while detecting clinically significant prostate cancer including cribriform growth with the ERSPC Rotterdam risk calculator and 4Kscore.* Urol Oncol. 2019;37:341-2.

6. **Verbeek JFM**, Nieboer D, Steyerberg EW, Roobol MJ. *Assessing a patient's individual risk of biopsy-detectable prostate cancer: Be aware of case mix heterogeneity and a-priori likelihood*. Eur Urol Oncol. 2019 Aug 17:S2588-9311(19)30115-4.

7. **Verbeek JFM**, Nieboer D, Parker C, Kattan MW, Steyerberg EW, Roobol MJ. *A tool for shared decision making on referral for prostate biopsy in the primary care setting: Integrating risks of cancer with life expectancy*. J Pers Med. 2019 Apr 22;9(2):19.

8. **Verbeek JFM**, Hannink G, Defoort KC, Wymenga AB, Heesterbeek PJC. *Age, gender, functional KSS, reason for revision and type of bone defect predict functional outcome 5 years after revision total knee arthroplasty: A multivariable prediction model.* Knee Surg Sports Traumatol Arthrosc. 2019;27:2289-96.

9. **Verbeek JFM**, Bangma CH, Kweldam CF, van der Kwast TH, Kummerlin IP, van Leender G, et al. *Reducing unnecessary biopsies while detecting clinically significant prostate cancer including cribriform growth with the ERSPC Rotterdam risk calculator and 4Kscore.* Urol Oncol. 2019;37:138-44.

10. Remmers S, **Verbeek JFM**, Nieboer D, van der Kwast T, Roobol MJ. *Predicting biochemical recurrence and prostate cancer-specific mortality after radical prostatectomy: Comparison of six prediction models in a cohort of patients with screening- and clinically detected prostate cancer.* BJU Int. 2019 May 4;124(4):635–42.

11. Alberts AR, Roobol MJ, **Verbeek JFM**, Schoots IG, Chiu PK, Osses DF, et al. *Prediction of high-grade prostate cancer following multiparametric magnetic resonance imaging: Improving the Rotterdam European Randomized Study of Screening for prostate cancer risk calculators.* Eur Urol. 2019;75:310-8.

12. **Verbeek JFM,** Roobol MJ. *What is an acceptable false negative rate in the detection of prostate cancer?* Transl Androl Urol. 2018;7:54-60.

13. Van Calster B*, Wynants L*, **Verbeek JFM**, Verbakel JY, Christodoulou E, Vickers AJ, et al. *Reporting and interpreting decision curve analysis: A guide for investigators.* Eur Urol. 2018;74:796-804.

14. Schoots IG, Osses DF, Drost FH, **Verbeek JFM**, Remmers S, van Leenders G, et al. *Reduction of MRI-targeted biopsies in men with low-risk prostate cancer on active surveillance by stratifying to PI-RADS and PSA-density with different thresholds for significant disease.* Transl Androl Urol. 2018;7:132-44.

15. Pereira-Azevedo N*, **Verbeek JFM***, Nieboer D, Bangma CH, Roobol MJ. *Head-to-head comparison of prostate cancer risk calculators predicting biopsy outcome.* Transl Androl Urol. 2018;7:18-26.

16. Mannaerts CK, Gayet M, **Verbeek JF**, Engelbrecht MRW, Savci-Heijink CD, Jager GJ, et al. *Prostate cancer risk assessment in biopsy-naive patients: The Rotterdam Prostate Cancer Risk Calculator in multiparametric magnetic resonance imaging-transrectal ultrasound (TRUS) fusion biopsy and systematic TRUS biopsy.* Eur Urol Oncol. 2018;1:109-17.

17. Roobol MJ*, **Verbeek JFM***, van der Kwast T, Kummerlin IP, Kweldam CF, van Leenders G. *Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer Risk Calculator for initial prostate biopsy by incorporating the 2014 International Society of Urological Pathology Gleason Grading and Cribriform growth.* Eur Urol. 2017;72:45-51.

18. **Verbeek JFM**, Roobol MJ, Steyerberg EW. *Risk based selection for active surveillance.* In: L Klotz. Active Surveillance for Localized Prostate Cancer, 2nd edt. Chapter 6. Springer, 2017.

19. Pereira-Azevedo N, Braga I, **Verbeek JF**, Osorio L, Cavadas V, Fraga A, et al. *Prospective evaluation on the effect of interobserver variability of digital rectal examination on the performance of the Rotterdam Prostate Cancer Risk Calculator*. Int J Urol. 2017;24:826-32.

20. **Verbeek J**, Hermans E, van Vugt A, Frölke JP. *Correct positioning of percutaneous iliosacral screws with computer-navigated versus fluoroscopically guided surgery in traumatic pelvic ring fractures.* J Orthop Trauma. 2016;30:331-5.

21. Frietman B, **Verbeek J**, Biert J, Frölke JP. *The effect of implant failure after symphyseal plating on functional outcome and general health.* J Orthop Trauma. 2016;30:336-9.

22. Verbeek AJ, **Verbeek JF**, van Dijck JA, Verbeek AL. *Predicting chance of disease: calculation using prediction rules.* Ned Tijdschr Geneeskd. 2014;158:A7041. [In Dutch]

* These authors contributed equally

LIST OF ABBREVIATIONS

AS	Active Surveillance
AUA	American Urological Association
AUC	Area Under the receiver operating characteristic Curve
Bx	Biopsy
CaP	Prostate Cancer
CI	Confidence Interval
C-index	Concordance statistic index
CR	Cribrifrom
csPCa	clinically significant Prostate Cancer
DCA	Decision Curve Analysis
DRE	Digital Rectal Examination
DUA	Dutch Urology Association
EAU	European Association of Urology
ERSPC	European Randomized Study of Screening for Prostate Cancer
FH	Family History
FN	False Negative
FP	False Positives
FPR	False Positive Rate
fPSA	free Prostatic Specific Antigen
GAP3	Global Action Plan Prostate Cancer Active Surveillance
GP	General Practitioner
GS	Gleason Score
HG	High Grade prostate cancer
HR	High Risk
IDC	Intraductal Carcinoma
IPA	Index of Prediction Accuracy. Refers also to India Pale Ale
IPSS	International Prostate Symptom Score
IQR	Interquartile Range
ISUP	International Society of Urological Pathology
LE	Life Expectancy
LR	Low Risk
mpMRI	multi-parametric Magnetic Resonance Imaging
NB	Net Benefit
NHG	Nederlands Huisartsen Genootschap
OR	Odds Ratio

PCa	Prostate Cancer
PCA3	Prostate Cancer Antigen 3
PCUK-RWG	Prostate Cancer United Kingdom Prostate Risk Working Group
PHI	Prostate Health Index
PIRADS	Prostate Imaging Reporting and Data System
PIVOT	Prostate Cancer Intervention versus Observation Trial
PLCO	American Prostate, Lung, Colorectal and Ovarian
PPV	Positive predictive value
PRIAS	Prostate cancer Research International: Active Surveillance
PSA	Prostatic Specific Antigen
PSAD	Prostatic Specific Antigen Density
PSMA-PET	Prostate-Specific Membrane Antigen-Positron Emission Tomography
PV	Prostate Volume
RC	Risk Calculator
ROC	Receiver Operating Characteristic curve
RPCRC	Rotterdam Prostate Cancer Risk Calculator
RR	Relative Risk / Risk Ratio
RU	Relative Utility
SEER	Surveillance, Epidemiology, and End Results
TP	True Positive
TPR	True Positive Rate
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
TRUS	Trans Rectal UltraSound
T-stage	Tumor stage
WW	Watchful Waiting
4Kscore	Four Kallirein Score

DANKWOORD

Dit proefschrift is een compilatie geworden van het werk en de input van vele mensen, die ik hierbij graag wil bedanken voor hun bijdragen. Met jullie steun is dit proefschrift tot een mooi eind gekomen. Rest nog de vraag: waarom lopen mensen weg bij de credits van een film, maar lezen ze wel het dankwoord van een thesis?

Mijn dank gaat allereerst uit naar mijn beide promotors. Prof. Dr. M.J. Roobol, beste Monique. Jij hebt me als PhD-student onder je wing genomen en je hebt me niet meer losgelaten. Jouw deur stond altijd voor mij open, je keek mijn stukken binnen de kortste keren na, iets waar andere PhD-studenten vaak alleen van konden dromen. Samen vormden we een goed team en konden vele onderzoeken neerzetten en publiceren. Jouw doorzettingsvermogen en zorg voor de mens op individueel en maatschappelijk niveau kan ik erg waarderen.

Prof. Dr. E.W. Steyerberg, beste Ewout. Na onze eerste ontmoeting had ik meteen door dat ik bij deze professor der professoren gelijk het PhD traject wilde ingaan. Met jouw hulp heb ik predictiemodellen mij eigen weten te maken. Ik heb zeer genoten van onze discussies zowel binnen als buiten de wetenschap om door de jaren heen. Ondanks de vele PhD studenten die je in meerdere universiteiten begeleidt, maakte je altijd tijd voor mij vrij. Ik heb bewondering voor de passie waarmee je onderzoek verricht en zal het leuk vinden om in de toekomst nog meer artikelen met je te publiceren.

Dear Prof. M.W. Kattan, dear Mike, Thank you for your kindness and the warm welcome in Cleveland, OH, USA. I am grateful for everything I learned from you in that period. Also, I'd like to thank you for your 'hypothetical rabbit' article which has formed the inspiration for the design of my thesis, a wonderful analogy with G. Crile's work in 1955.

Uiteraard ben ik ook verguld met alle co-auteurs die hebben bijgedragen aan mijn publicaties. Beste Prof. Dr. C.H. Bangma, dank voor uw vertrouwen in mij en de leuke discussies bij de journal club en op congressen. Prof. Dr. B. van Calster, de eerste keer dat ik je ontmoette was met een IPA-tje (jazeker) in Boston, beter kon niet. Bedankt voor de samenwerking en verdieping in terms of net benefit en decision curves, oorspronkelijk tot leven gebracht door Prof. dr. A.J. Vickers. Beste Prof. Dr. P.J.E. Bindels en Prof. Dr. S. Le Cessie, bedankt voor uw werk in de promotiecommissie; ik zie uit naar onze discussies tijdens mijn verdediging.

Frank-Jan Drost (Paranimf), mijn arts-onderzoeker maatje. Al vanaf het begin van mijn ErasmusMC periode was jij er en je hebt me door de jaren heen gesleept. Die talloze gesprekken over waar een threshold neer te leggen en wat dat wel niet moest betekenen voor de praktijk. Discussies die een dag konden duren, en zo nodig herhaald, tot frustratie van

onze kamergenoten en later zelfs tuingenoten. Succes met de afronding van je eigen proefschrift, de opleiding tot huisarts en natuurlijk jullie kleine, het komt voor elkaar.

Daniël Osses (Paranimf), je bracht alom gezelligheid met je mee! Altijd aanwezig en lekker aan het genieten van het vak urologie. Bedankt dat jij mijn paranimf wilt zijn op deze dag, nog amper een week voor je sollicitatie voor de opleiding tot uroloog. Succes: je wordt een topuroloog.

Dr. Alberts, Arnout, ik wil je bedanken voor je vriendelijkheid en bereidheid om iedereen te helpen waaronder mijzelf. Door jou voelde ik me gelijk welkom bij de urologie. Mocht ik iets aan mijn prostaat krijgen weet ik bij wie ik aanklop. Henk Luiting, bedankt dat ik zo af en toe bij je kon overnachten in Rotterdam en het geouwehoer op de afdeling, waardoor ik extra genoot van de rust tijdens mijn thuiswerkdagen. Verder wil ik mijn collega-onderzoekers bedanken voor de leuke tijd die we samen hebben meegemaakt (random volgorde seed 6841: Isabelle, Thomas, Sophie, Ilse, Toscane, Tess, Sebastiaan, Renée, Maaike, Leonard, Sarah, Nuno, Lisanne, Bodine, Kai, Michelle, Peter, Kim, Charlotte, Joep, Chris). Nog een ‘feitje van de dag’ voor jullie: dit proefschrift is nu af!

Daan Nieboer, bedankt voor je goede uitleg en mij de weg te wijzen in R codes en statistiek. Door jou ben ik een veel scherpere coder geworden. Ik waardeer je niet aflatende geduld ook al kwam ik af en toe last minute bij je aankloppen.

Beste Dr. Schoots, beste Ivo, bedankt voor de leuke samenwerking, tijdens de periodes van de prostaatbiopt poli en zeker ook tijdens congressen. Je had altijd leuke ideeën en ik genoot van onze gesprekken. Bedankt voor je support en begeleiding in het veld van de radiologie.

Onderzoek opzetten, uitvoeren, opschrijven en publiceren is een enorme onderneming wat duur en tijdrovend kan zijn waarin samenwerking centraal staat. Ik wil mijn dank uitspreken aan de onderzoekers, artsen en onderzoeksverpleegkundigen van het Erasmus MC, de European Randomized Study of Screening for Prostate Cancer (ERSPC) en de Movember Foundation het Global Action Plan 3 (GAP3). In het bijzonder Lionne, Conja, Marlies, Sophie, Jozien bedankt voor jullie input, advies en werving en opname van de studiepopulaties. Graag bedank ik ook de prostaatbiopten-verpleegkundigen Chris, Andrea en Daphne, die mijn dinsdagmiddagen of maandagochtenden altijd leuk maakten en ik weer naar de volgende bioptsessie uitkeek. Prof. F.H. Schröder, bedankt voor uw inzet voor de ERSPC-studie en al het prostaatwerk binnen de urologie. Ik vind het een eer u te hebben ontmoet. Om dit proefschrift te maken zijn in totaal van meer dan 300.000 patiënten gegevens gebruikt om zo tot de vele voorspellingen te komen. Ik ben alle deelnemers die deel hebben genomen in één of meerdere van de onderzoeken uit dit proefschrift enorm erkentelijk voor hun bijdrage.

Jan-Hendrik Venhuizen, één van mijn beste vrienden zo niet de beste. Ik ken je al vanaf het moment dat je kleiner was dan dat ik nu ben. Many thanks voor je geweldige Engelse zinnen, en vooral je mentale support tijdens dit PhD-traject. Top dat ik altijd op je kan vertrouwen, is ook wederzijds. Ik kijk uit naar jouw academische verdediging. Mijn vriend Manuel Kerssemakers, bedankt voor de epische 'The3Elements' momenten die we hebben meegemaakt en waarvan er nog vele zullen komen. Laurens Hendriks, ik zie je graag bij de volgende NBA-games ;). 'De groep': thnx voor jullie support, and till the next event!

Mijn Nijmeegse geneeskundevrienden en skigroep (Tom, Manon, Michiel, Jadeena, Diederik, Marit, Tim, Tijko, Eefke, Sabrina, Carine), bedankt dat jullie er zijn bij mijn verdediging en kan niet wachten met jullie weer op een berg te staan en naar beneden te vliegen. Tevens dank aan de pokerbazen (Jaron, Bas, Jordy, Michiel) en uiteraard ook mijn Spetters (Denise, Annemiek, Dipti, Marit en Niels), vrienden uit mijn intro en cogroep 151, begint het idee van een PhD-opleiding nu ook niet bij jullie te kriebelen?

The Love Squad/Pool Party (Victor, Pedro, Jordi, Simone, Miriam, Iris, Enna), ik ben later bij jullie gekomen, maar heb me nooit eerder ergens meteen zo welkom gevoeld. Ik kijk uit naar het volgende feestje. Speciale credits aan de lovesquad members aka "Who's the boss?" Elke en Dylan, voor jullie motiverende speeches en de leuke activiteiten die we hebben ondernomen. Speaking of the real boss: Pippa jij bent een geweldig dier die af en toe over het toetsenboard liep waardoor de meeste bizarre typo's ontstonden.

My international friends, especially Yi Hua, Gerardo and Phil. You showed me a different way of life and enriched my life, for which I will be forever grateful. I hope you are healthy, safe and happy, and hope to see you again after the corona crisis.

Roger Staats, English lecturer, thank you for your guidance, subtleties and corrections of the articles making them easily readable and even more informative.
Jane Klein, grafisch vormgever, plaatjes spreken meer dan woorden, reuze bedankt hiervoor, je bent een sterke ondernemer waar ik nog veel van kan leren.
Kim Nguyen, bedankt dat ik je opmaak van je mooie thesis heb mogen gebruiken, succes met je huisartsopleiding.

Beste Pa en Ma, André en Marian Verbeek, zonder jullie onvoorwaardelijke steun en vertrouwen was dit boek niet afgekomen. Ik ben zo blij dat ik jullie als ouders heb en jullie mij de vaste basis konden bieden om van daaruit verder te kunnen groeien. Anne (zus) en Niels, Frank (broertje) en Maartje: bedankt voor jullie steun in welk opzicht dan ook. Ik kon letterlijk bij jullie aankloppen ook al was het weer zo'n belachelijk verzoek van mijn kant. Ik kijk ernaar uit om oom te worden en hem of haar net als jullie wijze lessen over het leven bij te brengen.

De laatste persoon is altijd de belangrijkste persoon heb ik mij laten vertellen.
Lieve Lieke, mijn sweetie-petitie, jij noemt me altijd schildpadje. Niet omdat ik niet uit met schild durf te komen, maar omdat ik slow en steady ben. En oké ook omdat ik heel soms wat traag van begrip kan zijn. Maar gelukkig kan je nu toch inzien met deze schildpad metafoor dat het niet verkeerd is om een schildpadje te zijn. Liekje je bent mijn energieballetje, die mij soms helemaal gek kan maken, maar ook heel hard aan het lachen kan maken. Je geeft symbolisch kleur en betekenis aan mijn proefschrift zoals je dat aan mijn leven geeft.

PhD PORTFOLIO

Name	Johannes Franciscus Marcus Verbeek
PhD period	2016 – 2019
Erasmus MC department	Urology
Research school	NIHES – Netherlands Institute for Health Sciences
Promotors	Prof. M.J. Roobol-Bouts, Prof. E.W. Steyerberg

	Year	Workload (ECTS)
1. PhD training		
Courses		
Erasmus MC English Scientific writing Course	2016-2017	3.0
Erasmus MC Scientific Integrity Course	2016	0.3
Erasmus Summer Programme:	2016	1.4
ESP62 - Markers and Prediction Research		
ESP64 - Master class: Advances in Epidemiologic Analysis		
Statistical courses:		
Introduction in R (Datacamp)	2016-2019	10.0
Statistic modeling in R (Datacamp)		
Dealing with missing data in R (Datacamp)		
Visualization and imputation of missing data (Udemy)		
Advanced analytics with big data in R (Udemy)		
Advanced R programming (Pearson)		
Machine Learning in R (Packt)		
Advanced machine learning with R (Packt)		
Seminars and workshops		
Department of Urology Journal Club	2016-2019	1.0
Department of Urology Symposia	2016-2019	1.0
Presentations		
Annual EMUC-meeting, Milan	2016	1.0
European Association of Urology Annual Congress, London	2017	1.0
European Association of Urology Annual Congress, Copenhagen	2018	1.0
American Urological Association Annual Meeting, Boston	2017	1.0
Movember's Global Action Plan Prostate Cancer Active Surveillance Initiative Annual Meeting, Boston	2017	0.5
American Urological Association Annual Meeting, San Francisco	2018	1.0
European Randomized Study of Screening for Prostate cancer meeting, Rotterdam	2017	0.5

Conferences		
American Urological Association Annual Meetings	2017-2018	1.0
European Association of Urology Annual Congresses	2016-2019	1.5
European Multidisciplinary Congress on Urological Cancers Annual Meeting	2016	1.0
Annual meeting 'Werkgroep Epidemiologisch Onderzoek Nederland', Groningen	2019	0.5
Movember's Global Action Plan Prostate Cancer Active Surveillance Initiative Annual Meeting	2017, 2018	1.0
Other		
Cleveland Clinic, Research Scholar	2016	0.8
Secretary of Cause of Death Committee, ERSPC	2017-2019	1.0
Patient information day, SWOP, 50-year Urology	2019	0.3
2. Teaching		
Medical Students and Residents		
NIHES Winter course (Advanced Analysis of Prognosis Studies)	2017,2018	1.0
NIHES Summer course	2018	1.0
Statistics meta-analysis course, medical students	2017, 2018	1.0
3. Awards and awarded scholarships		
Best Poster Award Poster presentation: What is an acceptable false negative rate in the detection of prostate cancer? Annual AUA congress, San Francisco, US.	2018	
Best Poster Award Poster presentation: A prospective evaluation on the effect of inter-observer variability of DRE on the performance of the DRE based Rotterdam Prostate Cancer Risk Calculator. Annual AUA congress, Boston, US.	2017	
Reisbeurs René Vogels Stichting Work visit in Cleveland Clinic, Cleveland, US for the development and updating of international prostate cancer prediction models	2016	

The Hypothetical Rabbit

Michael W. Kattan

A common analogy for describing newly diagnosed prostate cancer patients is to think of turtles, rabbits, and birds. This is an analogy that prostate cancer appears to be attributable to Hinman, who borrowed from Crile when applied to breast cancer (1). Turtles are patients with very slow growing disease. Their disease grows so slowly that they need not be diagnosed, for the disease will never spread to the point of causing problems within the patient's lifetime. A turtle will die of another cause, not prostate cancer.

At the other extreme is the bird. The bird has been diagnosed too late to have impact on the disease. It has already spread and cannot be meaningfully slowed down, to the point where the patient is likely to die of his prostate cancer. For the opposite reason as the turtle, the bird is similarly not helped very much by a diagnosis of prostate cancer since it is already too late to stop the disease.

The rabbit sits in the sweet spot. The rabbit is the man with prostate cancer who needs to be diagnosed (his disease spreads faster than that of the turtle and indeed poses a threat to his life), yet the disease is still curable (unlike the disease borne of the bird).

This creature analogy is useful for thinking about prostate cancer screening. If you are a rabbit, it presumably makes sense for you to get screened. If you are a turtle or a bird, screening will potentially harm you (at least emotionally) and cannot help you, unless perhaps you are worried sick that you are a bird and would be relieved to find out you are a turtle.

It would seem that many patients who get screened and treated aggressively (say with surgery) believe they are rabbits. Many patients will thank a higher-level authority for having caught the cancer before it was too late, etc. This is natural, especially for emotional support. No one wants to go through an unnecessary surgical procedure, which is what happens when a turtle is operated on. Moreover, no one wants to go through a futile surgical procedure, which is what happens when a bird is operated on. The rabbit is at peace; he had a surgery that was necessary (i.e., he was not a turtle), and the surgery cured him (he is not a bird.) The desire to be labeled a rabbit is natural.

Presumably, most surgeons or radiation oncologists directly or indirectly convey to their treated patients that they were rabbits. No clinician really wants to admit that, unfortunately, the patient was treated unnecessarily (i.e., was indeed a turtle) or uselessly (i.e., was a bird). That is hardly satisfying for them, or their patient, which further motivates the popularity of the rabbit.

The real problem with the analogy, which is in widespread use, is that the rabbit is hypothetical. In real time, no patient actually knows if he is indeed a rabbit. Patients may believe they are/were rabbits; doctors may tell them they are/were rabbits. However, the truth is not revealed until the patient dies, and then only partially. This is unfortunate, clearly. A patient successfully treated surgically for his prostate cancer (i.e., is now alive and disease free) may indeed be a turtle. Once treated, it cannot be known with certainty what outcome the patient would have experienced had he not been treated. In addition, this same patient, apparently treated successfully with surgery, may tomorrow experience recurrence, and as such realize he is indeed a bird. This may happen at any point in the future, until death of the patient.

More bluntly, a patient diagnosed with prostate cancer yet left untreated until death from another cause was indeed a turtle. A patient diagnosed with prostate cancer and treated aggressively yet still succumbed to his prostate cancer was a bird. Any patient diagnosed with prostate cancer and alive cannot with certainty be classified as a turtle, rabbit, or bird. Once he dies, we will know if he was or was not a bird. The best we can do is to assign probabilities to each of these with statistical models that look at the nature of the disease, treatment received, age of the patient, his comorbidities, etc. But these are always going to be probabilities and, as such, hypotheticals, particularly for the rabbit.

1. Hinman F. Screening for prostatic carcinoma. J Urol (1991) 145:126–30.

Kattan MW. The Hypothetical Rabbit. Front Oncol. 2016 May 13;6:123.