Harms and Benefits of Prostate Cancer Screening and Active Surveillance

Gunstige en ongunstige effecten van het vroeg opsporing van prostaat kanker en actief opwachten

Tiago M. C. D. Marques

Erasmus University Rotterdam

Harms and Benefits of Prostate Cancer Screening

and Active Surveillance

Gunstige en ongunstige effecten van het vroeg opsporing van prostaat kanker en actief opwachten

Proefschrift

Ter verkrijging van de graad van doctor aan de

Erasmus Universiteit Rotterdam

Op gezag van de

rector magnificus

Prof.dr. H.A.P. Pols

En volgens de besluit van de College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 28 Juni 2017 om 1130

door

Tiago Manuel de Carvalho Delgado Marques

geboren te Lissabon, Portugal.

Erasmus University Rotterdam

Promotie Comissie

Promotor:	Prof.dr. Harry J. de Koning
Overige Leden:	Prof.dr. C. H. Bangma
	Prof.dr. T. M. De Reijke
	Dr. N. Pashayan
Co-Promotor:	Dr. E. A. M. Heijnsdijk

Table of Contents

Chapter One

Introduction

1.1 Prostate Cancer Epidemiology

Prostate cancer is the second most commonly diagnosed cancer and the 5th cause of cancer death worldwide (1). In the US it is the most diagnosed cancer for men, accounting for 26% of all cancers diagnosed and 9% of all cancer deaths in men (2). In 2015 in the US, an estimated 220,800 men were diagnosed from prostate cancer and about 27,540 men died from prostate cancer (2).

The highest incidence rates of prostate cancer are found in Australia/New Zealand, Northern America, Western and Northern Europe (3). In particular, Caribbean men, and African-Americans are at the highest risk of both prostate cancer detection and mortality (3, 4).

The incidence in these regions, and in particular the US, increased substantially in the early nineties due to the introduction of the Prostate Specific Antigen (PSA) test (Figure 1). On the other hand, prostate cancer mortality has experienced a downwards trend, possibly due to improvements in treatment and prostate cancer screening (5, Figure 2).

Figure 1: Age-adjusted prostate cancer incidence rates, per 100,000 men, for the US general population and African-Americans, in the period 1975-2012.

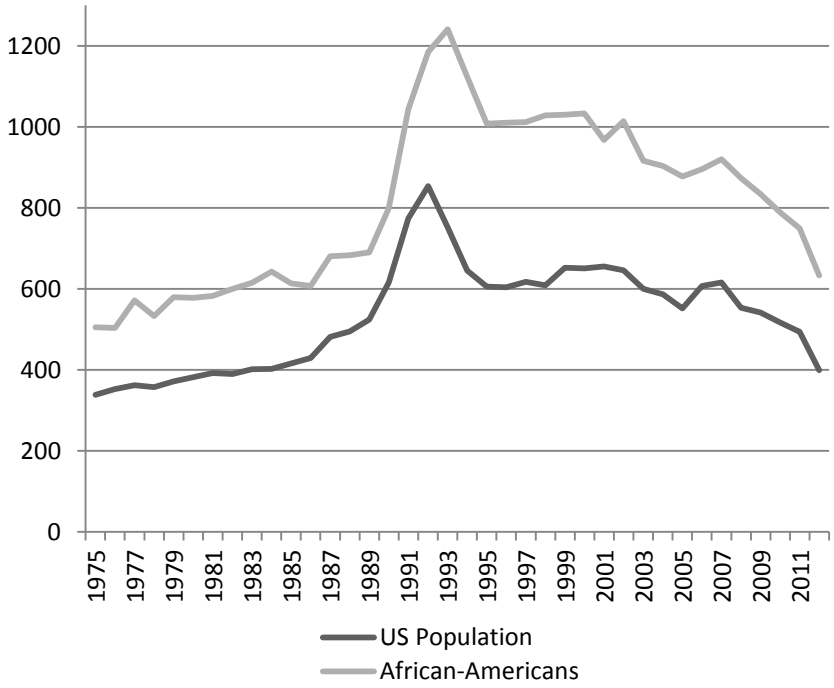
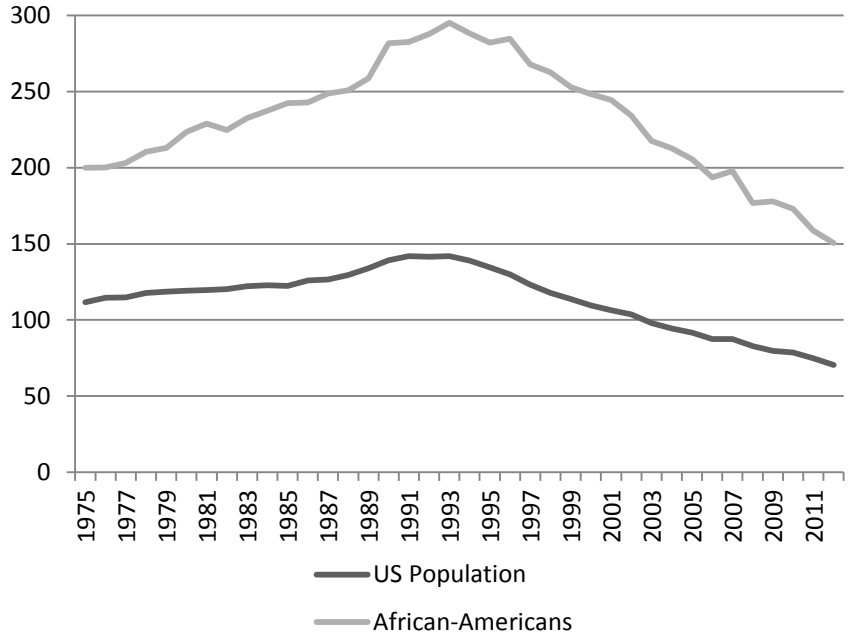


Figure 2: Age-adjusted prostate cancer mortality rates, per 100,000 men, for US general population and African-Americans, in the period 1975-2012.

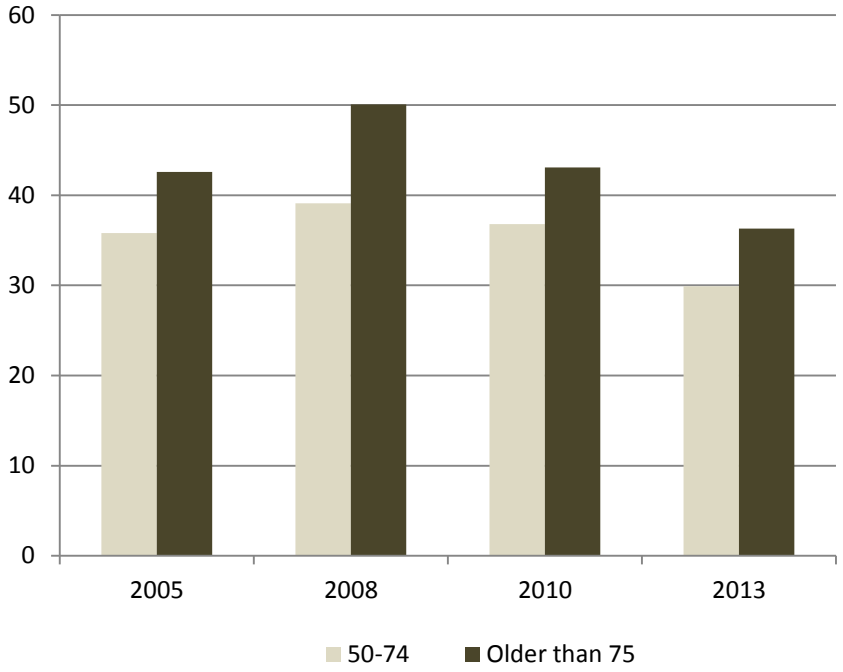


1.2 Prostate Cancer Screening

PSA is a protein produced in the prostate gland. The PSA-test consists of a simple blood test which gives the value of PSA in the blood. The higher the value is, the higher is the probability of cancer to be present, though an elevated PSA value could also be due to other causes. Usually, a PSA threshold of 3 or 4 is used for a positive screening test. After an elevated PSA, the patient will be referred for a prostate biopsy, and this will determine if the patient has cancer, and what is the severity of the disease. (6)

In the US in 2013, about 31% of men older than 50 years of age reported to have had a PSA-test in the last year. In practice, screening is performed in an inefficient way. For instance, while between 2008 and 2013, there was a significant decrease in the screening rate for men older than 75, which are unlikely to benefit from screening, however this rate is still higher than for men aged between 50 and 74 (Figure 3 and ref.7).

Figure 3: Usage of PSA-based screening in the US. *



* Based on (7). Some numbers are extrapolated from the published figure.

1.3 Evidence on the benefits of screening

The two biggest randomized clinical trials on prostate cancer screening were the ERSPC (European Randomized study of Screening for Prostate Cancer) (8) and the PLCO (Prostate, Lung, Colorectal and Ovarian cancer screening trial) (9). Other trials include the Norrkoping, Stockholm and Quebec studies, but a literature review concluded that they have a high risk of bias (10). Additionally, there is an ongoing trial on prostate cancer screening in the UK, CAP/ProtecT, which was originally designed as a trial for treatment of prostate cancer. This trial did not report any results yet about the efficacy of screening (11).

1.3.1 ERSPC trial

The ERSPC trial included 182,160 men, aged between 50 and 74, starting in 1994, in Belgium, Finland, France, Italy, The Netherlands, Spain, Sweden and Switzerland. The biggest centres in the study are Finland and the Netherlands. Men in the screening arm received PSA testing every 4 years (in Sweden every two years) between ages 50 and 74 (core age group 55-69), followed by prostate biopsy if the PSA value was 3 ng/ml or higher

in most centres. Men in the control arm received usual care. The latest report from the trial at 13 years follow-up showed a prostate cancer mortality reduction of 21% (8).

1.3.2 PLCO trial

The PLCO included 76,685 men aged between 55 and 74, in ten centres in the US, enrolled between 1993 and 2001. The screening arm applied annual PSA-testing during 6 years and digital rectal examination (DRE) for 4 years. It was recommended for men to undergo prostate biopsy if PSA $\geq$ 4 ng/ml. The latest report from the trial at 13 years of follow-up showed no mortality reduction due to prostate cancer screening (9). There were several reasons for this, namely the high degree of screening in the control arm (about 52%) and the low observed biopsy compliance (about 40%). (12, 13)

1.4 The harms of screening

While the blood draw for the PSA test is harmless, the subsequent prostate biopsy is not without harm. Several studies report a small risk of infection, and an increase in the risk of hospitalization compared to a control population (14-16). After screen-detection, a major concern is overdiagnosis and overtreatment. Overdiagnosis consists of the detection of cancer that would not have become clinically detected in absence of screening. If an overdiagnosed person is treated then we define the person as overtreated.

Main treatments for prostate cancer include surgery (radical prostatectomy) and radiation therapy. Radical prostatectomy results in the removal of the prostate gland. Radiation therapy directly kills cancer cells by using high-energy rays (17). Treatment for prostate cancer has major secondary effects, which are detrimental for the quality of life of the patient. A significant proportion of men treated by radical prostatectomy or radiation therapy report incontinence, impotence and bowel problems (18, 19).

Overdiagnosis is difficult to measure in a clinical trial. Ideally, we would need to follow all patients from randomization to death, to compute the true estimate of overdiagnosis, however we would need to wait decades for this. Estimates of overdiagnosis in prostate cancer can vary considerably depending on the method used or even between microsimulation models. For instance, three simulation models estimated a range for overdiagnosis between 23% and 42% of screen detected men (20). Another study using the excess incidence method, suggested that 60% of screen-detected men may be overdiagnosed (21).

In practice, these estimates are also difficult to translate from the population level to an individual patient. The probability that an individual benefits from screening and treatment can vary substantially depending on the individual person's characteristics. Namely,

overdiagnosis is highly age-dependent, with the majority of overdiagnosis cases occurring in men older than 60 (22-24). However, even within the same age group there could be large differences. For instance, Wever et al (24) estimated that for screen-detected men in the age group 65-69, the probability of overdiagnosis ranges from 9% to 50%, depending on disease stage at detection.

1.5 Active Surveillance

Active surveillance (AS) consists of the monitoring of men diagnosed with prostate cancer, but not yet treated, with PSA tests or repeat biopsies. The benefit of AS consists in the delay or avoidance of radical treatment, with the goal of helping patients to keep their pre-diagnosis quality of life. What is still unclear is whether the benefits of avoiding treatment outweigh the risk that a patient misses his cure by delaying radical treatment.

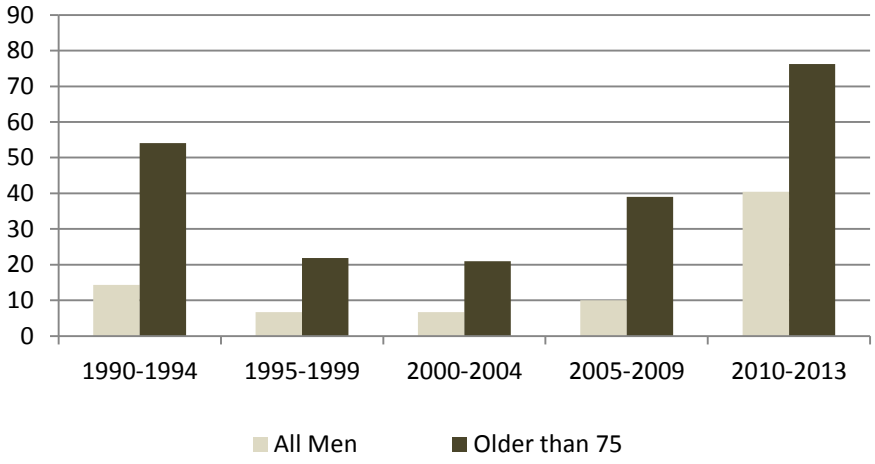
There are multiple AS cohorts designed to study the safety of this approach: Prostate Cancer Research International Active Surveillance (PRIAS) (25), UCSF cohort (26), Johns Hopkins (JH) (27) or the Toronto cohort (28) among others (29, 30). However, with the exception of the Toronto cohort, their median follow-up times are shorter than 5 years. So far, the results seem favorable, with few prostate cancer deaths and metastatic cases reported in their latest publications (25-28, 30), but there is a lack of long term results.

Most AS cohorts contain only one AS protocol. Usually, low-risk patients (stage <T2a and Gleason Score 6 or lower) are selected (30). While in some cohorts follow-up biopsies occur yearly (JH (27)), in others biopsy occurs up to every 3 years after the first year (PRIAS (25)). It is not yet clear whether AS is safe for intermediate risk patients. While in the UCSF cohort (26) it was found that the 4-year treatment-free survival (TFS) did not significantly differ between low and intermediate risk men, intermediate risk patients had significantly worse outcomes in the Toronto cohort (28). There are also differences regarding the criteria for referral to treatment. While in most cohorts grade and/or volume progression is the main criteria for referral to treatment (30), PSA velocity also plays a role in the Toronto cohort (28).

The usage of modelling to evaluate the lifetime outcomes of AS is necessary. Namely, there is a lack of long term results, since prostate cancer has a long lead time and its effects may be felt only years later and there are multiple ways of performing Active Surveillance, namely, different selection criteria or different time intervals between follow-up biopsies.

The current consensus is to offer AS for low-risk men, or at least to mention the option of AS to low-risk men (31). The uptake of Active Surveillance has been changing rapidly (32). For instance, in the CapSURE cohort, the usage of AS increased from less than 10% in 2010, to about 40% in 2013 (Figure 4, ref. 32). Across community practices in Michigan about half of the low-risk men receive AS (33).

Figure 4: Usage of Active Surveillance under low-risk men in the CapSURE cohort.*



* Based on (32). Including Watchful Waiting, particularly before 2000.

1.6 MISCAN: A microsimulation model for cancer screening

Here I give a brief description of the MISCAN model. For more technical details the reader is invited to read Chapter 12, Model Appendix. MISCAN (Microsimulation Screening Analysis) is a microsimulation semi-Markov model, where durations and transition probabilities to the next health state are dependent on current state, and where individual life histories are simulated. A man may develop prostate cancer depending on a probability of onset. If the person develops prostate cancer, then the time of onset is modelled based on a piecewise constant cumulative hazard, which depends on the particular age group. Each pre-clinical health state denotes a combination of T-stage, Gleason Score and absence or presence of metastasis. In each of these states, a person may become clinically diagnosed or screen-detected.

The screening module consists of a single joint sensitivity, which denotes sensitivity for PSA detection, biopsy compliance and biopsy sensitivity (See Chapters 3-5, Model Appendix). This was modified in Chapter 2, where we model PSA growth explicitly, and we separated biopsy sensitivity and biopsy compliance (See Chapters 2, 6-9, Model Appendix).

After detection, a patient is assigned to either radical prostatectomy (RP), radiation therapy (RT) or AS. The post-detection survival consists of three components, an effect of screening dependent on disease stage at detection (in Chapters 6-9, this changed to a continuous effect of screening depending on lead-time), an effect of treatment based on (34, 35) and a baseline survival, based on SEER data from the pre-PSA era.

For the European model, onset, transition probabilities and durations are calibrated to ERSPC data. For the US model, we calibrate the joint sensitivity of screening or PSA growth and an extra hazard of clinical diagnosis to SEER incidence data and using a generator which simulates an opportunistic screening schedule based on (36). Some of the PSA growth model parameters are based on literature (37), while others were freed to calibrate both to ERSPC's PSA distribution per age group and SEER incidence. The effect of screening is modelled as a cure probability which is calibrated to observed prostate cancer mortality reduction in the ERSPC (8).

1.7 Data

We populate the natural history model of MISCAN primarily with ERSPC data described above. In chapter 2, it was also used to model PSA growth. In order to adapt the model to a US population setting, the model was additionally calibrated to Surveillance, Epidemiology, and End Results (SEER) incidence data. SEER registry collects data on incidence and survival of multiple cancers throughout the US. It currently covers about 28% of the US Population (38).

In the Active Surveillance module of MISCAN, probabilities of referral to treatment given progression were calibrated to data from the Johns Hopkins cohort (27). It is a single arm observational study, where the main outcomes are overall and prostate specific survival. The Johns Hopkins cohort started recruiting men in 1995, and in the latest report it includes 1298 very low risk (Gleason 6, T1c and lower) and low-risk (Gleason 6, stage T2a or lower) men. Main criteria for referral to radical treatment included biopsy reclassification, either by volume or gleason progression. We used treatment free-survival as the target dataset for the calibration. Additionally, biopsy compliance during AS is based on observed biopsy compliance during the PRIAS study (25).

1.8 Aims and Research Questions in this thesis

1.8.1 Aim I: Screening

In the first part of this thesis we will use this microsimulation model to project the effects of multiple screening policies in the population. A limitation of using clinical trials to inform screening policies is that given the multitude of possible strategies to implement screening, no trial would have enough power to detect what is the most efficient screening policy. In most studies in this aim, we will try to find the set of screening policies that have the best balance between harms and benefits, usually by studying the trade-off between prostate cancer mortality reduction/life years gained and overdiagnosis or the trade-off between costs and QALYs gained.

Research Question 1: "Can we find a set of screening policies that significantly reduce the amount of overdiagnosis, while keeping most of the prostate cancer mortality reduction?"

Research Question 2: "Can prostate cancer screening be cost-effective?"

Research Question 3: "What is the influence of comorbid conditions in the harms and benefits of cancer screening?"

In the research question 4, we will try to explain disparities in prostate cancer incidence between African-Americans and the general US population.

Research Question 4: "Why is the prostate cancer incidence higher in African-Americans than in the general US population?"

1.8.2 Aim II: Active Surveillance

While the results of Active Surveillance seem promising, with few cases of prostate cancer mortality or metastasis, there is a lack of observed lifetime outcomes. Additionally, given the number of possible avenues for selecting and following men during AS, the use of modelling to evaluate the outcomes of AS protocols is necessary.

In the second part of this thesis we added a module to MISCAN to simulate several Active Surveillance protocols. We validated this module with Johns Hopkins data, and projected several combinations of Active Surveillance Protocols and screening policies.

As in Aim 1, our goal is to find the set of policies with the best balance of harms and benefits. We will do this from several viewpoints, namely, by estimating the reduction in

overtreatment due to AS and extra prostate cancer mortality or at its costs and effects compared to immediate treatment.

Research Question 5: “What is the prostate cancer mortality increase and overdiagnosis reduction associated with Active Surveillance, in comparison with Immediate Radical Treatment?”

Research Question 6: “Do personal characteristics matter when choosing between immediate radical treatment and active surveillance in an older age group (65+)?”

Research Question 7: “Is Active Surveillance more cost-effective than Immediate Radical Treatment?”

Research Question 8: “When should Active Surveillance for prostate cancer stop if no progression is detected?”

1.8.3 Aim III: Parameter Uncertainty in MISCAN

Results obtained with microsimulation models are subject to uncertainty. In this aim, we will investigate the effect of parameter uncertainty, that is, uncertainty in the outcomes due to uncertainty in parameter estimates, in overdiagnosis estimates. Since this task can be computationally intensive, we will perform this while trying to minimize the running time of MISCAN needed to obtain a confidence interval for the outcomes.

Research Question 9: “Can we evaluate parameter uncertainty in MISCAN in a more efficient manner?”

References

1. IARC. GLOBOCAN 2012, Estimated Incidence Mortality and Prevalence Worldwide. http://globocan.iarc.fr/Pages/fact_sheets_population.aspx . Acessed 31/01/2016.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin. 2015;65(1):5-29.

3. WCRF. Prostate Cancer Statistics. http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/prostate-cancer-statistics Acessed Jan 2016.

4. American Cancer Society. Cancer Facts & Figures for African-Americans. http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-036921.pdf . Acessed 19/02/2016

5. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118(23):5955-63.

6. NCI/NIH. http://www.cancer.gov/types/prostate/psa-fact-sheet Acessed 31/01/2016.

7. Jemal A, Fedewa SA, Ma J, et al. Prostate Cancer Incidence and PSA Testing Patterns in Relation to USPSTF Screening Recommendations. JAMA. 2015;314(19):2054-61.

8. 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(9959):2027-35.

9. Andriole GL, Crawford ED, Grubb RL 3rd, et al. Mortality Results after 13 years of Follow-up. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009; 360:1310-19.

10. Ilic D, O'Connor D, Green S, Wilt TJ. Screening for prostate cancer: an updated Cochrane systematic review. BJU Int. 2011 Mar;107(6):882-91.

11. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med. 2016;375(15):1415-1424.

12. Pinsky PF, Andriole GL, Kramer BS, et al; PLCO Project Team. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 2005;173:746-50.

13. Pinsky PF, Blacka A, Kramer BS, et al. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. Clin Trials. 2010;7(4):303-11.

14. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol 2013;64:876-92.

15. Loeb S, Carter HB, Berndt SI, et al. Is repeat prostate biopsy associated with a greater risk of hospitalization? Data from SEER-Medicare. J Urol. 2013;189:867-70.

16. Nam RK, Saskin R, Lee Y, et al. Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. J Urol 2010;183:963-8.

17. American Cancer Society. http://www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-treating-general-info. Acessed 17/02/2016.

18. Korfage IJ, Essink-Bot ML, Borsboom GJ, et al. Five-year follow-up of health-related quality of life after primary treatment of localized prostate cancer. Int J Cancer. 2005;116(2):291-6.

19. Sanda MG, Dunn RL, Michalski J, et al. Quality of Life and Satisfaction with Outcome among Prostate-Cancer Survivors. N Engl J Med. 2008 Mar 20;358(12):1250-61.

20. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. J Natl Cancer Inst 2009; 101:374-83.

21. Welch HG, Black WC. Overdiagnosis in Cancer. J Natl Cancer Inst 2010;102:605–13.

22. Gulati R, Inoue LY, Gore JL, Katcher J, et al. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst 2014;106:djt367.

23. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC Med 2014;12:26. doi: 10.1186/1741-7015-12-26.

24. Wever EM, Hugosson J, Heijnsdijk EA, et al. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. Br J Cancer 2012;107:778-84.

25. Bul M, Zhu X, Rannikko A, et al. Radical Prostatectomy for Low-Risk Prostate Cancer Following Initial Active Surveillance: Results From a Prospective Observational Study. Eur Urol 2012;62:195-200.

26. Cooperberg M, Cowan J, Hilton J, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 2011;29:228-34.

27. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

28. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 2015;33:272-7.

29. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol 2012;62:976-83.

30. Simpkin AJ, Tilling K, Martin RM, et al. Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 2015;67:993-1005.

31. Klotz L. Active Surveillance for Prostate Cancer: Debate over the Application, Not the Concept. Eur Urol. 2015;67(6):1006-8.

32. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA 2015;314:80-2.

33. Womble PR, Montie JE, Ye Z, et al. Michigan Urological Surgery Improvement Collaborative. Contemporary Use of Initial Active Surveillance Among Men in Michigan with Low-risk Prostate Cancer. Eur Urol 2015; 67:44-50.

34. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118:5955-63.

35. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.

36. Mariotto AB, Etzioni R, Krapcho M, Feuer EJ. Reconstructing PSA Testing Patterns Between Black and White Men in the US from Medicare Claims and the National Health Interview Survey. Cancer. 2007;109(9):1877-86.

37. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.

38. Overview of the SEER Program. https://seer.cancer.gov/about/overview.html Acessed 15/05/2017.

39. 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 Oct 20;33(30):3379-85.

Chapter Two

Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk,
Harry de Koning.

Published: Int J Cancer. 2015;136(7):1600-7

Abstract

While the benefit of prostate specific antigen (PSA) based screening is uncertain, a significant proportion of screen-detected cases is overdiagnosed. In order to make screening worthwhile, it is necessary to find policies that minimize overdiagnosis, without significantly increasing prostate cancer mortality (PCM).

Using a microsimulation model (MISCAN) we project the outcomes of 83 screening policies in the US population, with different start and stop ages, screening frequencies, strategies where the PSA value changes the screening frequency, and strategies in which the PSA threshold (PSAt) increases with age.

In the base case strategy, yearly screening 50-74 with a PSAt of 3, the lifetime risk of PCM and overdiagnosis equals respectively, 2.4% and 3.8%.

The policies that reduce overdiagnosis the most (for maximum PCM increases relative to basecase of 1%, 3% and 5%, respectively) are with a PSAt of 3, (1) yearly screening 50-74 where, if PSA <1 at age 65 or older, frequency becomes 4 years, with 3.6% (5.9% reduction), (2) 2-year screening 50-72, with 2.9% (24.3% reduction) and (3) yearly screening 50-70 (PSAt of 4 after age 66), with 2.2% (43.4% reduction).

Stopping screening at age 70 is a reasonable way to reduce the harms and keep the benefit. Decreasing the stopping age has a larger effect on overdiagnosis reduction than reducing the screen frequency. Screening policies where the frequency of screening depends on PSA result or in which the PSAt changes with age did not substantially improve the balance of harms and benefits relative to simple yearly screening.

Introduction

There is not yet a consensus about the magnitude of the net benefit of early detection of prostate cancer, since the two largest prostate cancer screening randomized controlled trials found conflicting results. In the Prostate, Lung, Colorectal and Ovarian (PLCO) screening trial, after 13 years of follow-up, it was found that annual PSA screening does not reduce prostate cancer mortality.[1] On the other hand, the European Randomized Study of Screening for Prostate Cancer (ERPSC) trial, after a 11-year follow-up and with mainly a 4-year screening interval, found a 20% prostate cancer mortality (PCM) reduction in comparison with no screening.[2]

Current guidelines in the US for prostate cancer screening reflect the lack of consensus, on what is the best trade-off between harms and benefits: the U.S. Preventive Services Task Force (USPSTF) recommended in 2012, against PSA screening, while the American Urological Association (AUA) advised in 2013 that men should be screened between the ages 55 and 69. [3] Even within the organizations that are favourable to screening, different screening algorithms are proposed.[4]

In general, screening could result in some lives saved, with many overdiagnosed cases. Overdiagnosis is the detection of cancer, where in the absence of screening, the cancers would have never been detected. There is a small chance of complications due to biopsy after an elevated PSA[5]. However there are significant side effects associated with treatment that have a negative effect in the quality of life [6]. Overdiagnosis estimates can vary considerably depending on the methods and populations used [7], with number needed to detect to prevent one prostate cancer death between 5 and 48, and proportion of overdiagnosis in screen-detected cases varying between 22 and 67% [7-8].

In order to make PSA screening worthwhile, given the uncertainty surrounding the magnitude of the benefit of screening, and the estimates suggesting that a significant proportion of screen-detected cases is overdiagnosed, there is a clear need to find screening policies with a better balance of harms and benefits. We aim to find the combination of start-stop age, screening frequency and prostate-specific antigen threshold for biopsy referral (PSAt) that minimizes overdiagnosis for several thresholds of limited PCM increase compared to screening yearly with a PSAt of 3 between ages 50 and 74.

Previously several studies [6,9] found that stopping screening at an earlier age than 70 is more cost-effective than screening until age 74. Others [10-11] found that strategies where frequency is determined by PSA result or where a higher PSAt for older are

used could have a better balance of harms and benefits.

We add to Gulati et al [11] by evaluating more combinations of frequencies and stopping ages, and strategies in which the PSAt increases for older age groups. A promising new way of performing prostate cancer screening is to risk stratify men based on PSA level at a certain age [12,13]. Men with a lower PSA would then be screened less frequently. What is still unclear is whether this set of policies is effectively better than simple screening algorithms, and therefore we also compare the harms and benefits of several ways of doing PSA-based risk stratification.

Few people advocate screening beyond age 75. However, Berghdal et al.[14] noticed that after 10 years of stopping screening in the ERSPCtrial the incidence of high risk cancers in the screening group closely resembles the control (no screen) group. Therefore we also quantify the harms and benefits of screening for these older men.

Methods

Simulation Model

Microsimulation Screening Analysis (MISCAN) is a microsimulation model, which simulates individual life-histories. A detailed description is available in http://cisnet.cancer.gov/prostate/profiles.html . Each health state denotes a different disease, detection and treatment phase. There are three detection phases, preclinical, clinical and screen detected. We model 18 disease stages, consisting of the combination of 3 stages (T1, T2, T3), 3 grades (which correspond to Gleason Score 2-6, 7 and > 7) and whether or not the cancer is metastasized. In each of these disease stages, an individual can progress to a higher disease state, be clinically or screen detected, or can die.

The transition probabilities and durations, including the transitions between different stages, grades and duration until metastasized cancer are calibrated to ERSPC data. The durations between different disease states and between disease states and clinical detection are modelled as Weibull distributions with its parameters depending on T-stage and Gleason Score [15-16]. Relative to our ERSPC model, there is an extra hazard of clinical detection, which implies an earlier time of clinical diagnosis in the US [17].

After detection, the person is assigned to either watchful waiting, radical prostatectomy (RP) or radiation therapy (RT) with equal chance. A baseline prostate cancer survival curve is assigned, based on SEER data (1983-1986) from the pre-PSA era. The effect of treatment is introduced with a lower hazard ratio for PCM (0.62 for RP and 0.70 for RT) based on Etzioni et al [18]. Additionally there is a cure

rate based on the mortality reduction observed in the ERSPC trial and dependent on whether Gleason Score is lower or equal than 7 (0.42) or higher than 7 (0.23) [19].

A new feature in MISCAN, is the PSA growth generator. It resembles Inoue et al. [20], but with some important modifications (Model Appendix, Chapter 11). The PSA growth parameters together with biopsy sensitivities are calibrated to SEER incidence data from 1990 to 2002 and to the PSA distribution of the first round of the ERSPC screening trial, except the correlation between errors in the PSA growth equation, which was calibrated to ERSPC data from different screening rounds to predict the probability that PSA< 1 about age 60 and PSA > 3 two screening rounds later.

The simulated population of ten million is based on the lifetables of Mariotto et al.[21]. US screening patterns are imposed, based on the screening generator of Mariotto et al. [21]. Additionally we assume 100% biopsy compliance and 90% screen attendance in the basecase analyses.

Up to the year 2012, we reproduce the screening patterns in the US population. Afterwards a screening policy is implemented. We accept new cohorts of men aged 50, between 2012 and 2022. Additionally, men opportunistically screened prior to 2012 and not yet diagnosed are also screened. This allows us to incorporate the effect of current screening practices in our projections.

The main outcomes are the lifetime risk of prostate cancer mortality (PCM) and overdiagnosis. Here we define overdiagnosis as a person who is screen-detected with prostate cancer, but would not be diagnosed in the absence of screening and dies from other causes.

Screening Policies

The base case screening strategy is yearly PSA testing between ages 50 and 74 with a PSAt of 3. We compute combinations of screening policies with starting ages 50, 54, 58 and 62, and stop ages 62, 66, 68, 70, 72 and 74. The frequencies used are yearly, 2-year and 4-year. In addition, we investigated yearly screening until age 80 (Table 1).

Vickers et al.[12] and Roobol et al.[13] suggest that using PSA to do risk stratification might help to improve outcomes of standard screening policies. Therefore, for a set of yearly screening policies, we compute the effects of changing the screening frequency based on the PSA value at a certain age (60, 65 and 70). Additionally, we also study whether using a higher PSAt for older age groups, might help to improve the harms and benefits trade-off.

Table 1: List of screening policies by age, frequency, PSA threshold and PSA value or age dependent condition.

Policy	Start/Stop Age	Frequencies	PSAt*	Condition
Basecase				
1	50-74	yearly	3	
Varying Start/Stop Age and Screen Frequencies				
2	50-62	2-,4-year, yearly	3	
3	50-66	2-,4-year, yearly	3	
4	50-68	2-year, yearly	3	
5	50-70	2-,4-year, yearly	3	
6	50-72	2-year, yearly	3	
7	50-74	2-,4-year	3	
8	54-62	2-,4-year, yearly	3	
9	54-66	2-,4-year, yearly	3	
10	54-68	2-year, yearly	3	
11	54-70	2-,4-year, yearly	3	
12	54-72	2-year, yearly	3	
13	54-74	2-,4-year, yearly	3	
14	58-66	2-,4-year, yearly	3	
15	58-68	2-year, yearly	3	
16	58-70	2-,4-year, yearly	3	
17	58-72	2-year, yearly	3	
18	58-74	2-,4-year, yearly	3	
19	62-70	2-,4-year, yearly	3	
20	62-74	2-,4-year, yearly	3	
Higher PSAt by age				
21	50-70/72/74	Yearly	4	
22	50-70/72/74	Yearly	3	if $66 \leq$ age < 70, then PSAt = 4, if age $\geq$ 70 then PSAt = 5
23	50-70/72/74	Yearly	4	if $66 \leq$ age < 70, then PSAt = 5, if age $\geq$ 70 then PSAt = 7
Screen Frequency and Stop Age dependent on PSA result (PSA based risk stratification)				
24	50-74	yearly	3	if age ≥ 60 and PSA < 1 then Stop
25	50-74	yearly	3	if age ≥ 60 and PSA < 1 then frequency is 8-year
26	50-74	yearly	3	if age ≥ 60 and PSA < 1 then frequency is 4-year
27	50-74	yearly	3	if age ≥ 60 and PSA < 1 then frequency is 2-year
28	50-74	yearly	3	if age ≥ 65 and PSA < 1 then Stop
29	50-74	yearly	3	if age ≥ 65 and PSA < 1 then frequency is 4-year
30	50-74	yearly	3	if age ≥ 65 and PSA < 1 then frequency is 2-year
31	50-74	yearly	3	if age ≥ 70 and PSA < 1 then Stop

Screening Men Older than 75				
32	50-76/78/80	yearly	3	
33	50-76/78/80	yearly	3	if age ≥ 70 then PSAt = 4
34	50-76/78/80	yearly	3	if 66 ≤ age < 70, then PSAt = 4, if age ≥ 70 then PSAt = 5
35	50-76/78/80	yearly	4	if 66 ≤ age < 70, then PSAt = 5, if age ≥ 70 then PSAt = 7
AUA 2013 guideline based policies				
36	55-69	yearly	4	
37	55-69	2-year, yearly	3	

Finally, we investigate strategies based on the 2013 AUA guideline. This guideline recommends to screen men between ages 55 and 69. As the recommended PSAt and screening frequency is not clear, we run this policy for PSAt's 3 and 4 ng/ml, yearly and every 2 years.[3]

We perform sensitivity analyses by running several scenarios for some of the best policies (PCM increase thresholds: -3%, 0%, 3% and 6%). We run four scenarios assuming a reduced screening efficacy: a screening attendance of 70% and 50%, a decrease of 20% in biopsy sensitivity and a 41% biopsy compliance combined with a 85% attendance as observed in the PLCO trial [22]. We also vary the parameters that determine PSA growth by 20%, to get a range of possible screening outcomes.

Results

Model Validation

In Model Appendix Figure 2, we compare the predicted incidence by MISCAN compared to the observed incidence in the US SEER data between 1975 and 2009. The model reproduces the peak in prostate cancer incidence due to the introduction of PSA testing, but with slight overprediction. Beyond the calibration period, the predicted incidence is close to the observed incidence rate.

In Model Appendix Table 3, the PSA distribution projected by MISCAN is close to the PSA distribution observed in the 1st round of the ERSPC trial. The maximum difference between predicted and observed is about 6 percentual points (For more details see Model Appendix PSA Growth Generator Section).

Model Projections

In table 2, we show the outcomes of the reference strategy compared to stop screening in 2012. In the reference strategy the lifetime risk of prostate cancer detection is 14.0%. The lifetime risk of metastasis and PCM is, respectively, 0.7% and 2.4%, while if we stop screening after 2012, it becomes respectively, 1.7% and 2.8%. The PCM reduction due to screening is 19.4%, and the lifetime risk of overdiagnosis is 3.8% corresponding to 43.8% of the screen-detected cases.

Table 2: Basecase screening policy compared with stopping screening 2012.

Lifetime Risk %	50-74,1y, PSAt=3	Stop Screening 2012
Prostate Cancer Mortality	2.4	2.8
Metastasis	0.7	1.7
Overdiagnosis	3.8	0
Detection	14.0	9.7
Performance		
PCM Benefit %	19.4	-
Number Needed to Screen	162	-
Number Needed to Treat	15	-
Overdiagnosis (% of Screen Detected)	43.8	-
Metastasis (% of Detected)	4.7	17.8

* In the absence of screening (also prior to 2012) the lifetime risk of PCM is 3.0%.
# PSAt stands for prostate-specific antigen threshold for biopsy referral.

In table 3 we show the lifetime risk of overdiagnosis and PCM, divided by several thresholds of PCM increase, due to reduced screening intensity, and ordered by overdiagnosis reduction. Increasing the start age of screening from 50 to 54, 58 and 62 has little effect on overdiagnosis. Starting screening at age 54 seems to be the best age of these four, as starting earlier has little effect on PCM (0.2% increase from age 50), while starting later increases PCM (0.8% for age 58 and 2.2% for age 62, in Table 3).

Decreasing the stop age from 74 to an earlier age gives a profound effect. With yearly screening, stopping at 72, 70 or 68 reduces the lifetime risk of overdiagnosis, respectively, from 3.8% to 3.1% (18.0% reduction), 2.5% (34.0% reduction) and 2.0% (48.3% reduction). Correspondingly, PCM increases, due to less intensive screening, from 2.4% to 2.4% (1.6% increase), 2.5% (3.2% increase) and 2.5% (5.1% increase) only. Stopping screening at an earlier age than 68 can reduce overdiagnosis by more than 60%, but then PCM increases more than 6% in comparison with screening until age 74 (Table 3).

Keeping the stop age fixed, a reduction in the screening frequency from yearly to every 2 years or 4 years substantially reduces the amount of overdiagnosis (Table 3). However, in most cases, the effect of decreasing the stop age is more powerful than reducing the screen frequency. For instance, consider the policy where screen every two years between ages 50 and 72. The PCM increases from 2.4% to 2.5% (3.0% increase) and overdiagnosis decreases from 3.8% to 2.9% (24.3% reduction), while screening yearly between 50 and 70 increases PCM slightly more to 2.5% (3.2% reduction) but reduces overdiagnosis to 2.5% (34.0% reduction). Also see PCM increase thresholds 2%, 4%, 5% and 7% in Table 3.

Compared to the reference strategy, the three AUA inspired strategies (55-69 yearly with PSAt of 3 and PSAt of 4, 2-year with a PSAt of 3, respectively) reduce overdiagnosis from 3.8% to 2.2% (41.0% reduction), 1.9% (50.4% reduction) and 2.1% (45.3% reduction). PCM increases, respectively, from 2.4% to 2.5% (4.4% increase), 2.5% (6.3% reduction) and 2.5% (5.4% increase) (Table 3).

We also projected outcomes for policies where the frequency of screening changes according to the PSA value (Table 1). Stopping screening at ages 60 or 65 if the PSA is smaller than 1, reduces overdiagnosis from 3.8% to 2.8% (27.6% reduction) and 3.2% (17.1% reduction) but PCM increases from 2.4% to 2.5% (3.6% increase) and 2.4% (1.9% increase), respectively. This is a lower overdiagnosis reduction than other policies with similar PCM levels (Table 3).

A possible way to reduce overdiagnosis is to increase the PSAt (Table 1). The combination where we screen from age 50 with a PSAt of 3, increase to 4 after age 66, and 5 after age 70 has similar harms and benefits trade-off, as simple yearly screening strategy (Table 3). By contrast, most screening policies with an initial PSAt of 4 have a lower overdiagnosis reduction compared to other screening policies with similar PCM levels (See PCM increase thresholds 3%, 5%, 6% and 7% in Table 3).

Screening until age 76 or 80 decreases PCM from 2.4% to 2.3% (1.4% reduction) and 2.3% (3.6% reduction), respectively, and less metastasized cases. But this comes at a price of an overdiagnosis increase from 3.8% to 4.6% (20% increase) and 6.3% (64.6% increase), respectively (Table 3).

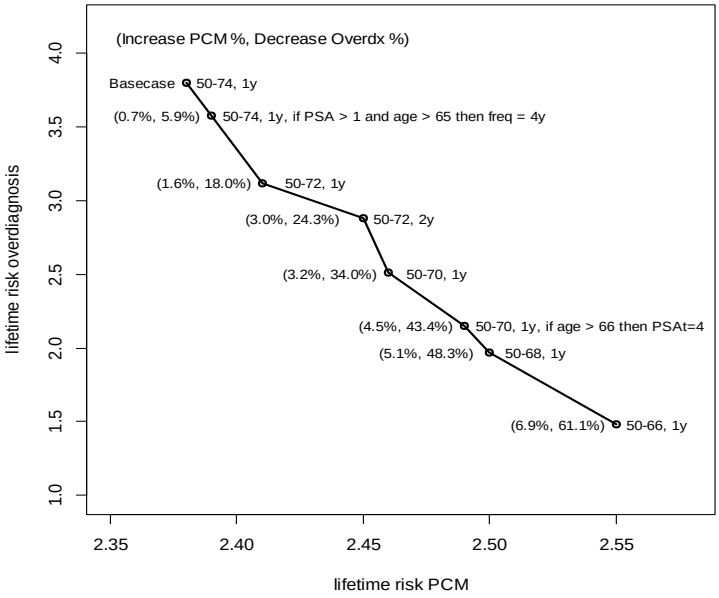
The policies that reduce overdiagnosis the most, for maximum PCM increases of 1%, 2%, 3%, 4% and 5% are, respectively, yearly screening 50-74 where, if PSA <1 at age 65 or older, screening frequency changes to 4 years, yearly screening 50-72, 2-year screening 50-72, yearly screening 50-70 (all with a PSAt of 3) and yearly screening 50-70 with a PSAt of 3 for men younger than 66 and PSAt of 4 for men older than 66 (Figure 1 and Table 3).

Table 4: Selected screening policies by PCM increase threshold, due to less intensive screening, relative to basecase (50-74, 1y, PSAt=3) and ordered by overdiagnosis reduction.

Maximum PCM Increase Threshold	Screening Policy	Probability Overdx (- Overdx %)	Probability PCM (+ PCM %)
0%	*50-74, 1y, PSAt = 3*	*3.80 (0.0)*	*2.38 (0.0)*
1%	50-74, 1y, PSAt = 3, if age > 65, PSAt < 1 then freq. = 4y	3.58 (5.9)	2.39 (0.7)
	58-74, 1y, PSAt = 3	3.81 (-0.1)	2.38 (0.8)
	54-74, 1y, PSAt = 3	3.81 (-0.2)	2.40 (0.2)
2%	50-72, 1y, PSAt = 3	3.12 (18.0)	2.41 (1.6)
	50-74, 1y, PSAt = 3, if age > 65, PSAt < 1 then stop	3.15 (17.1)	2.42 (1.9)
	50-74, 2y, PSAt = 3	3.51 (7.6)	2.41 (1.6)
3%	50-72, 2y, PSAt = 3	2.88 (24.3)	2.45 (3.0)
	50-74, 1y, PSAt = 3, if 66 ≤ age < 70 then PSAt = 4, if age ≥ 70 then PSAt = 5	3.10 (18.6)	2.43 (2.2)
	50-74, 1y, PSAt = 4	3.31 (12.8)	2.44 (2.6)
	62-74, 1y, PSAt = 3	3.76 (1.1)	2.43 (2.2)
4%	50-70, 1y, PSAt = 3	2.51 (34.0)	2.46 (3.2)
	50-74, 1y, PSAt = 3, if age > 60, PSAt < 1 then stop	2.75 (27.6)	2.47 (3.6)
	50-74, 4y, PSAt = 3	3.07 (19.4)	2.47 (3.8)
5%	50-70, 1y, PSAt = 3, if age > 66 then PSAt = 4	2.15 (43.4)	2.49 (4.5)
	55-69, 1y, PSAt = 3	2.24 (41.0)	2.48 (4.4)
	50-70, 2y, PSAt = 3	2.31 (39.2)	2.49 (4.5)
	50-74, 1y, PSAt = 4, 66 ≤ age < 70 then PSAt = 5, if age ≥ 70 then PSAt = 7	2.61 (31.2)	2.49 (4.6)
6%	50-68, 1y, PSAt = 3	1.97 (48.3)	2.50 (5.1)
	55-69, 2y, PSAt = 3	2.08 (45.3)	2.51 (5.4)
	50-70, 1y, PSAt = 4	2.13 (43.9)	2.51 (5.5)
7%	50-66, 1y, PSAt = 3	1.48 (61.1)	2.55 (6.9)
	50-70, 1y, PSAt = 4, 66 ≤ age < 70 then PSAt = 5, if age ≥ 70 then PSAt = 7	1.84 (51.6)	2.54 (6.4)
	55-69, 1y, PSAt = 4	1.88 (50.4)	2.54 (6.3)
	50-70, 4y, PSAt = 3	1.99 (47.5)	2.53 (6.3)
17%	*Stop Screening 2012*	*0.00 (100.0)*	*2.83 (16.2)*
Screening Men Older than 75			
-1%	50-76, 1y, PSAt = 3	4.56 (-20.0)	2.34 (-1.4)
-3%	50-80, 1y, PSAt = 3	6.26 (-64.6)	2.30 (-3.6)

All other screening policies are shown in Online Supplement tables 3 (by screening frequency and start stop age) and 4 (PSA based risk stratification, PSAt dependent on age, screening men older than 75).
Reduction (-) or increase (+), in percentage, relative to basecase (50-74, PSAt=3, yearly screening). Screening policies are ordered by overdiagnosis reduction.
* PSAt stands for prostate-specific antigen threshold for biopsy referral, Overdx for Overdiagnosis, PCM for prostate cancer mortality.

Figure 2: The screening policies that reduce the overdiagnosis the most, for each prostate cancer mortality (PCM) threshold.



*All policies with a PSA threshold for biopsy referral of 3, unless stated otherwise. The numbers between brackets are relative to the basecase screening policy (50-74, 1y, PSAt=3).
PSAt stands for prostate-specific antigen threshold for biopsy referral, Overdx for Overdiagnosis, PCM for prostate cancer mortality.

Sensitivity Analyses

Varying the PSA growth parameters by 20%, lowering the sensitivity or the attendance does not seem to have a major impact in the results, resulting in the worst case in a slightly lower screening benefit. However, if we apply together the biopsy compliance and attendance observed in the PLCO trial, there is a significant reduction in screening benefit. The lifetime risk of PCM for 50-74 yearly screening at a PSAt of 3 is about 2.5%, which is close to stopping screening at age 68, under our initial assumption of 100% biopsy compliance. On the other hand, the risk of overdiagnosis is about 2.9% stopping at 74, with PLCO compliance compared to 2.0% with 68 as stop age and with 100% biopsy compliance. (Supplementary Information Table 3 and 4).

Discussion

For the reference strategy of screening yearly between 50 and 74 with a PSAt of 3, we predict a lifetime risk of PCM of 2.4% (19.4% PCM reduction due to screening) and a corresponding 3.8% lifetime risk of overdiagnosis (43.8% of screen-detected cases).

Overdiagnosis estimates can diverge considerably, depending on the methods and population used [7-8]. Based on ERSPC, Welch et al.[23] estimated 60% of screen-detected cases are overdiagnosed, while Draisma et al [17], used 3 simulation models and estimated a range between 23 and 42%. Gulati et al.[11] found a lifetime risk of overdiagnosis and PCM of 3.3% and 2.2%, respectively, for a similar strategy of yearly screening between 50 and 74 with PSAt 4, which compares to 3.3% and 2.4%, in our model. Wu et al [24] estimated a 3.4% lifetime risk of overdiagnosis in Finland, for men screened between 55 and 67 every 4 years.

Ranking screening policies by harms and benefits is a difficult task. Usually we cannot reduce PCM without increasing overdiagnosis. Additionally, the weighting of harms and benefits is highly subjective. However, one could intuitively say, that if a screening strategy causes a small decrease in the benefit and a large decrease in the harms, one would prefer such a policy. Following this principle and according to our model's predictions, it seems hard to justify screening for prostate cancer in average men older than 70. Namely, we predict that if one screens yearly from age 50 to 70, overdiagnosis reduces from 3.8% to 2.5% (34.0% reduction) and PCM from 2.4% to 2.5% (3.2% increase), compared to screening until 74.

It should be stressed however, that this conclusion is dependent on the particular individual's health and life expectancy. Additionally, our model represents an "ideal" screening environment, which contrasts with the 41% biopsy compliance observed in PLCO, and that more accurately represents the US reality [22]. Screening yearly, with the biopsy compliance as in the PLCO between ages 50 and 74 gives a similar PCM level, as if we would have 100% biopsy compliance and stop at age 68. The overdiagnosis level of the latter though, is much lower, suggesting there could be a large efficiency gain by screening less, but enforcing a significantly higher biopsy compliance. In practice though, biopsy compliance is likely not random as in MISCAN. It could be dependent on life expectancy or variables which are also predictors of the severity of the disease like PSA value.

If we compare policies with similar levels of PCM, it seems that the effect of decreasing the stopping age on overdiagnosis reduction is larger than reducing the screening frequency or increasing the initial PSAt to 4. This can seem somewhat counterintuitive, as prostate cancer is a slow growing disease, but one should also

notice that the probability of overdiagnosis is highly age dependent [25-27].

We also evaluated screening policies inspired by the 2013 AUA guideline. We find that screening between 55 and 69 is a reasonable way to go forward, as with later ages overdiagnosis increases substantially, and there is a very small difference in mortality between starting screening at age 50 and age 54. However, our predictions are more favourable for yearly screening with a PSA threshold for biopsy referral of 3, than screening every two years or with a PSAt of 4.

PSA based risk stratification reduces overdiagnosis. However, it seems that for a similar PCM increase, the overdiagnosis reduction is smaller than the effect of decreasing the stopping age. This could be the case since few men with PSA smaller than 1 at age 60 will develop the disease in our model. This is in accordance with the ERSPC trial, where the probability of having a PSA lower than 1 at age 60 and a PSA higher than 3, in the next two rounds is only 2.3% (Online Supplement Table 2). A limitation of these analyses is that we did not vary the stop age. Also other ways of doing PSA based risk stratification could be considered.

Screening men older than 75 saves some lives, but each additional year of screening adds a large number of overdiagnosed cases, about 20% for additional 2 yearly screening rounds, and it is therefore undesirable, at least for a person with average life expectancy.

Due to the high number of parameters in MISCAN it is not possible to do a (probabilistic) sensitivity analysis including all parameters and assumptions used to construct the model could have an influence in the projections. For instance, the results from Gulati et al.[11] do not differ much from our own, but we verify that for similar screening strategies, the overdiagnosis level is similar but PCM is lower, thus their model is slightly more favourable to screening.

In this study we also did not model Active Surveillance (AS). This type of treatment can reduce the harms of screening, namely by delaying and in some cases avoiding the side-effects associated with radical treatment. On the other hand, its long term effects still needs to be investigated as most current AS observational cohorts have a small follow-up [28]. Referring screened men to AS who are at low and perhaps intermediate risk could greatly reduce the harms of screening. The inclusion of quality of life and costs associated with each disease state can potentially change the relationship between harms and benefits. For instance, Heijnsdijk et al.[6] found that adding quality of life estimates reduced the benefit of screening by 23%. Extrapolating that result to this study could mean that stopping screening earlier would be preferred. Also introducing costs could potentially favour stopping screening earlier.

In general we should screen men with a life expectancy higher than the expected duration from onset of the disease to a state where the prostate cancer is not curable. In most cases the patient's health status will give an idea about the life expectancy, but since the onset of the disease is not observed and there is considerable variation in its duration, it remains a difficult task, to predict which men may benefit of screening.

Our results support the view that screening men older than 70 yearly and indiscriminately can lead to a large increase of overdiagnosis. However, depending on the maximum tolerated level of PCM, earlier stopping ages could be considered. Additionally, we find that screening at lower frequencies than yearly and newly proposed screening protocols in which the frequency depends on PSA result [12,13] or where the PSAt increases with age, do not seem to improve much on simple yearly screening.

References

1. Andriole GL, Crawford ED, Grubb, et al. Mortality Results after 13 years of Follow-up. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009; 360:1310-19.

2. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med. 2012;366:981-90.

3. Carter HB, Albertsen PC, Barry MJ, et al. Early Detection of prostate cancer: AUA guideline. American Urological Association 2013. http://www.auanet.org/education/guidelines/prostate-cancer-detection.cfm , Accessed February 2014.

4. Roobol, MJ, Carlsson SV. Risk stratification in prostate cancer screening. Nat. Rev. Urol. 2013;10:38-48.

5. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol. 2013;64:876-92.

6. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality of Life Effects of Prostate Specific Antigen Screening. N Engl J Med. 2012;367:595-605.

7. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. . Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med. 2013;158:831-8.

8. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and Overtreatment of Prostate Cancer. Eur Urol. 2014;65:1046-55.

9. Nichol MB, Wu J, Huang J,et al. Cost-effectiveness of Prostate Health Index for prostate cancer detection. BJU Int. 2007;110:353-62.

10. Kobayashi T, Goto R, Ito K, Mitsumori K. Prostate cancer screening strategies with re-screening interval determined by individual baseline prostate-specific antigen values are cost-effective. EJSO. 2007;33:783-89.

11. Gulati R, Gore JL and Etzioni R. Comparative Effectiveness of Alternative Prostate-Specific Antigen–Based Prostate Cancer Screening Strategies. Ann Intern Med. 2013; 158:145-53.

12. Vickers AJ, Lilja H. Predicting prostate cancer many years before diagnosis: how and why?. World J Urol. 2012;30:131-35.

13. Roobol MJ, Roobol DW, Schroder FH. Is additional testing necessary in men with prostate-specific antigen levels of 1.0 ng/mL or less in population-based screening setting?. Urology. 2005;65:343-46.

14. Grenabo Bergdahl A, Holmberg E, Moss S, Hugosson J. Incidence of Prostate Cancer After Termination of Screening in a Population-based Randomised Screening Trial. Eur Urol. 2013;64:703-09.

15. Draisma G, Boer R, Otto SJ, et al. Lead time and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003;95:868-78.

16. Draisma G, Postma R, Schröder FH, et al. Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer. 2006;119:2366-71.

17. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. J Natl Cancer Inst. 2009; 101:374-83.

18. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118:5955-63.

19. Wever EM, Draisma G, Heijnsdijk EA, de Koning HJ. How does early detection by screening affect disease progression?: Modelling estimated benefits in prostate cancer screening. Med Decis Making. 2011;31:550-58.

20. Inoue LY, Etzioni R, Morrell C, Müller P. Modeling Disease Progression with Longitudinal Markers. J Am Stat Assoc. 2008;103:259-70.

21. Mariotto AB, Etzioni R, Krapcho M, Feuer EJ. Reconstructing PSA Testing Patterns Between Black and White Men in the US from Medicare Claims and the National Health Interview Survey. Cancer. 2007;109(9):1877-86.

22. Pinsky PF, Andriole GL, Kramer BS, et al PLCO Project Team. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 2005;173:746-50.

23. Welch HG, Black WC. Overdiagnosis in Cancer. J Natl Cancer Inst 2010;102:605–13.

24. Wu GH, Auvinen A, Määttänen L, et al. Number of screens for overdetection as an indicator of absolute risk of overdiagnosis in prostate cancer screening. Int J Cancer. 2012;131:1367-75.

25. Gulati R, Inoue LY, Gore JL, et al. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst. 2014;106.

26. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC Med. 2014 Feb 11;12:26. doi: 10.1186/1741-7015-12-26.

27. Pashayan N, Duffy SW, Pharoah P, et al. Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. Br J Cancer. 2009;100:1198-204.

28. Cooperberg MR, Carroll PR, and Klotz L. Active Surveillance for Prostate Cancer: Progress and Promise. J Clin Oncol. 2011;29:3669-76.

29. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.

Supplementary Material

Additional Outcomes

Table 1: Screening policies per start-stop age and screening frequency, with PSAt of 3

Reduction (-) or increase (+), in percentage, relative to basecase (50-74, PSAt=3, yearly screening)

Yearly Screening

Strategy	Age	Freq.	PSAt	- Overdx %	+ PCM %	+ Mx %
Stop Screen 2012	-	-	-	*100.0*	*16.2*	*62.2*
1	*50-74*	*Yearly*	*3*	*0.0*	*0.0*	*0.0*
2	50-62	Yearly	3	80.8	10.5	50.8
1y	50-66	Yearly	3	61.1	6.9	39.9
1y	50-68	Yearly	3	48.3	5.1	32.6
1y	50-70	Yearly	3	34.0	3.2	23.3
1y	50-72	Yearly	3	18.0	1.6	12.6
1y	54-62	Yearly	3	80.7	10.7	51.0
1y	54-66	Yearly	3	60.9	7.1	40.1
1y	54-68	Yearly	3	48.1	5.3	32.8
1y	54-70	Yearly	3	33.8	3.5	23.7
1y	54-72	Yearly	3	17.8	1.8	12.9
1y	54-74	Yearly	3	-0.2	0.2	0.7
1y	58-66	Yearly	3	61.2	7.6	40.8
1y	58-68	Yearly	3	48.3	5.8	33.7
1y	58-70	Yearly	3	34.0	4.0	24.9
1y	58-72	Yearly	3	17.9	2.4	14.5
1y	58-74	Yearly	3	-0.1	0.8	2.8
1y	62-70	Yearly	3	35.2	5.3	28.1
1y	62-74	Yearly	3	1.1	2.2	8.1
1y	*55-69*	*Yearly*	*3*	*41.0*	*4.4*	*28.6*

Screening Every 2 years

Strategy	Age	Frequency	PSAt	- Overdx %	+ PCM %	+ Mx %
2	50-62	2-year	3	82.6	11.1	51.8
3	50-66	2-year	3	64.5	7.8	42.0
4	50-68	2-year	3	52.6	6.2	35.3
5	50-70	2-year	3	39.2	4.5	27.2
6	50-72	2-year	3	24.3	3.0	17.7
7	50-74	2-year	3	7.6	1.6	7.1
8	54-62	2-year	3	82.2	11.1	51.8
9	54-66	2-year	3	63.9	7.9	41.9
10	54-68	2-year	3	51.9	6.2	35.3
11	54-70	2-year	3	38.6	4.6	27.2
12	54-72	2-year	3	23.6	3.0	17.8
13	54-74	2-year	3	6.7	1.6	7.2

14	58-66	2-year	3	63.7	8.1	42.3
15	58-68	2-year	3	51.6	6.4	35.8
16	58-70	2-year	3	38.2	4.8	27.8
17	58-72	2-year	3	23.2	3.3	18.7
18	58-74	2-year	3	6.2	1.9	8.4
19	62-70	2-year	3	38.8	6.1	30.5
20	62-74	2-year	3	6.8	3.3	12.7
37	*55-69*	*2-year*	*3*	*50.4*	*6.3*	*34.6*

Screening Every 4 years						
Strategy	Age	Frequency	PSAt	- Overdx %	+ PCM %	+ Mx %
2	50-62	4-year	3	85.2	11.9	53.3
3	50-66	4-year	3	69.6	9.1	45.2
5	50-70	4-year	3	47.5	6.3	33.5
7	50-74	4-year	3	19.4	3.8	18.1
8	54-62	4-year	3	84.3	11.8	52.9
9	54-66	4-year	3	68.3	9.0	44.7
11	54-70	4-year	3	45.9	6.2	32.7
13	54-74	4-year	3	17.6	3.8	17.4
14	58-66	4-year	3	67.3	9.1	44.6
16	58-70	4-year	3	44.7	6.5	32.7
18	58-74	4-year	3	16.1	4.1	17.7
19	62-70	4-year	3	44.4	7.3	34.1
20	62-74	4-year	3	15.6	4.9	20.2

* PSAt stands for prostate-specific antigen threshold for biopsy referral, Overdx for Overdiagnosis, PCM for prostate cancer mortality and Mx for Metastasis.
** See Appendix Table 9 for all 2 and 4-year runs and the note on Cohort Effects.
*** In italic, screening policies based on the 2013 AUA guideline.

Table 2: Higher PSAt per age, PSA based risk stratification and Screening men older than 75

Reduction (-) or increase (+), in percentage, relative to basecase (50-74, PSAt=3, yearly screening)						
Higher PSAt's per age						
Strategy	Age	Freq.	PSAt	- Overdx %	+ PCM %	+ Mx %
21	50-70	Yearly	4	43.9	5.5	30.3
21	50-72	Yearly	4	29.3	4.0	21.7
21	50-74	Yearly	4	12.8	2.6	11.7
22	50-70	Yearly	3*	43.4	4.5	28.9
22	50-72	Yearly	3*	33.4	3.5	22.9
22	50-74	Yearly	3*	18.6	2.2	14.0
23	50-70	Yearly	4*	51.6	6.4	34.1
23	50-72	Yearly	4*	44.1	5.7	30.3
23	50-74	Yearly	4*	31.2	4.6	23.4
36	*55-69*	*Yearly*	*4***	*45.3*	*5.4*	*31.5*

*In this policy the PSAt depends on age:
22: if 66 ⩽ age < 70, then PSAt = 4, if age ⩾ 70 then PSAt = 5.
23: if 66 ⩽ age < 70, then PSAt = 5, if age ⩾ 70 then PSAt = 7.
** 36: screening policy based on the 2013 AUA guideline.

PSA based risk stratification (Screen Frequency and Stop Age dependent on PSA result)

Initial Screening Strategy: 50-74, 1y, PSAt=3

Strategy: PSA based condition	- Overdx %	+ PCM %	+ Mx %
24: If PSA<1 from age 60 then stop screening.	27.6	3.6	26.2
25: If PSA<1 from age 60 then 8 year frequency.	18.0	2.3	17.2
26: If PSA<1 from age 60 then 4 year frequency.	9.4	1.3	9.3
27: If PSA<1 from age 60 then 2 year frequency.	3.5	0.5	3.5
28: If PSA<1 from age 65 then stop screening.	17.1	1.9	15.7
29: If PSA<1 from age 65 then 4 year frequency.	5.9	0.7	5.6
30: If PSA<1 from age 65 then 2 year frequency.	3.6	0.4	3.2
31: If PSA<1 from age 70 then stop screening.	2.9	0.3	2.3

Screening Men Older than 75

Strategy	Age	Freq.	PSAt	- Overdx %	+ PCM %	+ Mx %
32	50-76	Yearly	3	-20.0	-1.4	-13.8
32	50-78	Yearly	3	-41.9	-2.6	-27.9
32	50-80	Yearly	3	-64.6	-3.6	-39.7
33	50-76	Yearly	3*	-5.3	1.4	1.3
33	50-78	Yearly	3*	-25.3	0.3	-9.2
33	50-80	Yearly	3*	-46.3	-0.5	-17.9
34	50-76	Yearly	3*	1.9	1.1	4.6
34	50-78	Yearly	3*	-16.6	0.1	-4.9
34	50-80	Yearly	3*	-36.1	-0.7	-13.2
35	50-76	Yearly	4*	16.4	3.6	16.2
35	50-78	Yearly	4*	-0.1	2.8	9.0
35	50-80	Yearly	4*	-17.7	2.1	3.0

*In this policy the PSAt depends on age:
33: if age ≥ 70, then PSAt = 4.
34: if 66 ≤ age < 70, then PSAt = 4, if age ≥ 70 then PSAt = 5.
35: if 66 ≤ age < 70, then PSAt = 5, if age ≥ 70 then PSAt = 7.

PSAt stands for prostate-specific antigen threshold for biopsy referral, Overdx for Overdiagnosis, PCM for prostate cancer mortality and Mx for Metastasis.

Sensitivity Analyses

Table 3: Basecase screening policy (50-74, PSAt=3, yearly screen), attendance, biopsy compliance and sensitivity

	Attendance			PLCO Biopsy Compliance	Sensitivity	PSA growth	
Lifetime Risk %	90%	70%	50%	41%	80%	-20%	20%
Overdiagnosis	3.8	3.6	3.3	2.9	3.6	3.3	4.3
PCM	2.4	2.4	2.4	2.5	2.4	2.4	2.3
Metastases	0.7	0.7	0.7	0.8	0.7	0.8	0.6
Detection	14.0	16.7	16.2	13.4	16.8	13.6	14.5
Performance							
Mortality Benefit %	19.4	18.8	17.6	15.3	18.6	17.3	21.4
NNS	161.7	166.1	175.3	205.1	168.5	181.9	146.2
NNT	15.2	15.0	14.8	15.7	15.3	15.3	15.2
Overdiagnosis (% of Detected)	43.8	43.6	43.7	40.4	43.2	42.3	45.2
Mx (% of Detected)	4.7	5.5	5.0	6.2	4.9	5.6	3.8

* Basecase 90% attendance and 100% biopsy compliance. For the PLCO biopsy compliance, all the sensitivities were multiplied 0.41 and with an attendance of 85%, as observed on the PLCO trial ref. For the sensitivities, all sensitivity parameters were multiplied by 0.8. For the "PSA growth", parameters b_{1i} and b_{2i}, described in table 6, are multiplied by 0.8 or 1.2.

** PCM stands for prostate cancer mortality, NNS stands for number needed to screen to save one life, NNT for number needed to treat to save one life and Mx for Metastasis.

Table 4: Sensitivity Analyses

Reduction (-) or increase (+), in percentage, relative to basecase (50-74, PSAt=3, yearly screen)			
Different screening efficacy assumptions			
Age, Frequency	- Overdx %	+ PCM %	+ Mx %
Attendance 50%			
50-68, yearly	49.5	4.6	29.4
50-72, 2-year	31.6	3.9	21.5
50-80, yearly	-66.5	-3.2	-33.6
Attendance 70%			
50-68, yearly	48.8	4.9	31.4
50-72, 2-year	27.3	3.5	19.4
50-80, yearly	-65.4	-3.5	-37.6
PLCO Biopsy Compliance and Attendance			
50-68, yearly	54.6	4.2	26.9
50-72, 2-year	39.7	4.2	23.2
50-80, yearly	-72.4	-3.0	-29.4
Biopsy Sensitivity*80%			
50-68, yearly	49.3	4.9	31.7
50-72, 2-year	27.3	3.3	19.4
50-80, yearly	-65.9	-3.5	-37.9
Varying PSA growth parameters *			
Age, Frequency	- Overdx %	+ PCM %	+ Mx %
-20%			
50-68, yearly	49.8	4.4	27.7
50-72, 2-year	25.8	2.5	15.4
50-80, yearly	-64.7	-3.0	-28.6
+20%			
50-68, yearly	47.2	5.7	37.6
50-72, 2-year	24.1	3.3	21.2
50-80, yearly	-63.7	-4.2	-51.7

* PSA growth equation parameters b_{1i} and b_{2i} , described in table 6, are multiplied by 0.8 or 1.2 .

Policies selected based on the best screening policies for maximum mortality increase thresholds -3%, 0%, 3% and 6% . All Policies with a PSAt of 3.

Overdx stands for Overdiagnosis, PCM for prostate cancer mortality and Mx for Metastasis.

Chapter Three

Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data

EAM Heijnsdijk, TM de Carvalho, A Auvinen, M Zappa, V Nelen, M Kwiatkowski, *et al.*

Published: J Natl Cancer Inst. 2014;107(1):366

Abstract

Background: The results of the ERSPC trial showed a statistically significant 29% prostate cancer mortality reduction for the men screened in the intervention arm and a 23% negative impact on the life-years gained due to quality of life. However, alternative prostate-specific antigen (PSA)-screening strategies for the population may exist, optimizing the effects on mortality reduction, quality of life, overdiagnosis and costs.

Methods: Based on data of the ERSPC trial, we predicted the numbers of prostate cancers diagnosed, prostate cancer deaths averted, life-years and quality-adjusted life-years (QALY) gained and cost-effectiveness of 68 screening strategies starting at age 55, with a PSA threshold of 3, using micro-simulation modeling. The screening strategies varied by age to stop screening and screening interval (1 to 14 years or once in a lifetime screens) and therefore number of tests.

Results: Screening at short intervals of 3 years or less was more cost-effective than using longer intervals. Screening at ages 55-59 with 2-year intervals had an incremental cost-effectiveness ratio of $73,000 per QALY gained and was considered optimal. With this strategy, lifetime prostate cancer mortality reduction was predicted as 13%, and 33% of the screen-detected cancers were overdiagnosed. When better quality of life for the post-treatment period could be achieved, an older age of 65-71 years for ending screening was obtained.

Conclusion: Prostate cancer screening can be cost-effective, when it is limited to two or three screens between ages 55-59 years. Screening above age 63 years is less cost-effective due to loss of QALYs because of overdiagnosis.

Introduction

The European Randomised study of Screening for Prostate Cancer (ERSPC) has shown a disease-specific mortality reduction of Prostate-Specific-Antigen (PSA) screening for prostate cancer [1]. After eleven years of follow-up, prostate cancer mortality was reduced by 29% after adjustment for noncompliance. In terms of absolute effect, 37 cancers would need to be detected to avert one prostate cancer death [1]. Some of the screen-detected prostate tumors (23% to 42%) might never give rise to clinical symptoms and would not lead to death from prostate cancer [2].

These overdetected cancers reduce quality of life and result in higher costs due to overtreatment [3], affecting the balance of benefits and harms as well as cost-effectiveness of PSA testing for prostate cancer. In our recent study, we could demonstrate that the introduction of a screening program between the ages of 55-70 with a four-year interval would result in a gain of 52 life-years and 41 quality-adjusted life-years (QALYs) per 1000 men over their life span [4].

Very recently the AUA recommended shared decision-making for men age 55 to 69 years that are considering PSA screening, but they gave no clear indication of the screen interval. In the ERSPC, the Swedish center used a 2-year screening intervals, whereas the other centers used 4-year intervals [1]. In the United States, annual screening is more common. There are no trials comparing different screening intervals and such empirical studies are highly unlikely to be conducted because of immense resources required.

Few recent cost-effectiveness studies have been published using QALYs gained. Most cost-effectiveness studies for prostate cancer screening have been performed before large screening trial results had been published and showed very inconsistent results [5, 6].

The aim of present study is to assess the cost-effectiveness of prostate cancer screening. Based on data of the ERSPC trial various prostate cancer screening strategies were modeled to find the optimal screening intervals and ages.

Methods

The MISCAN model

MISCAN, MIcrosimulation SCreening ANalysis, was used for the evaluation of prostate cancer screening. The MISCAN prostate cancer model was developed in 2003 [7]. Since, the model was adjusted to explicitly model the metastatic stages,

treatment, survival and cure rates [8, 9].

MISCAN is a stochastic model that simulates individual life histories. The natural history of prostate cancer starts with a transition from "no prostate cancer" into preclinical screen-detectable prostate cancer. Tumor development is modeled as a progression through 18 stages (a combination of clinical T-stage T1, T2 and T3+, differentiation grade Gleason sum less than 7, 7 and more than 7 and metastatic stage 0 or 1). In each preclinical stage, the tumor may progress into another preclinical stage, become screen detected or clinically diagnosed (Model Appendix Figure 1). For each individual, the model predicts two life histories: one in the absence of screening and one in the presence of screening.

The cancers are divided into clinically diagnosed cancers, relevant screen-detected cancers and overdetected cancers (cancers that would not have become clinically diagnosed during a person's life). The model parameters for the disease and the test sensitivity are estimated with the use of data from the Rotterdam and Göteborg ERSPC centers (46,000 men, age 55-69), and the Dutch National Cancer Registry and the model is validated with the use of incidence data of all ERSPC centers. Other cause mortality is modelled using the Dutch life expectancy. The model and validation have been described before [4].

The treatment assignment in MISCAN is based on age, stage and Gleason score specific distribution of primary treatments (radiation therapy, radical prostatectomy and active surveillance) in the Rotterdam center of the ERSPC. It is assumed that 30% of men under active surveillance receive a secondary treatment within 7 years. All men dying of prostate cancer as well as all men with metastases receive palliative treatment.

Survival without treatment was modelled by using the Gleason score-specific survival curves for men detected with locoregional prostate cancer [10]. For distant disease, survival curves based on SEER data were used. The effects of treatment were modelled by assuming a relative risk of dying of 0.65 for radical prostatectomy [11] compared with watchful waiting. The same relative risk was assumed for radiation therapy.
The cure rate assumption is used to calculate the survival: a proportion of the screen-detected men with a local regional cancer will be cured and the remaining are not cured and die of prostate cancer or other causes at exactly the same time as they would have in a situation without screening. This stage-dependent cure rate was estimated for a prostate cancer mortality reduction of 29% after a follow-up of 11 years for men who attended at least one screen, corresponding to the prostate cancer mortality reduction of screened men in the ERSPC [1]. This resulted in cure rates of 0.51 for Gleason Score less than 7, 0.30 for Gleason Score 7 and 0.11 for

Gleason Score more than 7.

Screening Protocols

A cohort of 10 million men aged 55 in 2012 was simulated. Screening programs started in 2012, with 80% participation at each round. Screening intervals of 1, 2, 3, 4, 6, 8, 10, 12 and 14 years, starting at age 55 were simulated as well as a once in a lifetime screens. The age at which screening was stopped was varied between 55 and 75. The corresponding costs and effects were calculated until the year 2060 when all men in the cohort had died.

Costs

The unit costs of screening, diagnoses, primary treatment, follow-up and palliative care were obtained from literature [3]. The costs were calculated in dollars by using the purchase power parity for health [12]. Indirect costs were not included. The number of screening visits, diagnoses, prostate cancer deaths, treatments and life-years were predicted by the MISCAN model. To take into account biopsies with a negative result, the total number of biopsies was calculated by using the number of screen detected cancers and a mean positive predictive value of 22.7% of a biopsy in the screen arm of the ERSPC [13] and by using the number of clinically detected cancers and the positive predictive value of 35.8% of a biopsy in the control arm [14].

Quality of Life

QALYs were calculated by using utility estimates, values between 0 (death or worst imaginable health) and 1 (full health), representing patient desirability of a particular health state. Utility estimates and durations of all phases in screening, diagnoses and treatment of prostate cancer were obtained from literature (Table 1) [4]. The loss in QALYs was calculated by multiplying the loss in utility with the duration of the phase in Table 1 and the number of men in a phase obtained from MISCAN. For example when 800 men are screened once, they lose 800 x 0.01 x 1/52 year = 0.15 QALYs due to the screening itself.

Table 1. Costs, utility estimates and durations of the various phases in screening, diagnosis and treatment, obtained from previous studies [3, 4]*

Intervention	Unit costs ($)	Health state	Utility estimates (range)	Duration
Screening	39	Screening Attendance	0.99 (0.99-1)	1 week
Invitation	3.2			
Blood Sample	15.5			
PSA determination	20.3			
Diagnosis	277	Diagnostic Phase	0.90 (0.87-0.94)	3 weeks
Biopsy	150	Diagnosis	0.80 (0.75-0.85)	1 month
PA-Research	54			
GP-Consulting	73			
Primary Therapy and Follow-up				
Staging	326			
RP	19,235	RP (< 2 mo)	0.67 (0.56-0.9)	2 months
RT	23,110	RP (> 2 mo)	0.77 (0.70-0.91)	10 months
AS	2,588			
19 PSA tests	680	RT (< 2 mo)	0.73 (0.71-0.91)	2 months
10 DRE	800	RT (> 2 mo)	0.78 (0.61-0.88)	10 months
4 biopsies	1108	AS	0.97 (0.85-1.00)	Max. 7 years
Follow-up	245	One year after treatment	0.95 (0.93-1.00)	9 years
Advanced Disease				
Palliative Therapy	20,000	Palliative therapy	0.60 (0.24-0.86)	30 months
		Terminal illness	0.40 (0.24-0.40)	6 months

* AS denotes Active Surveillance, RT denotes Radiation Therapy and RP denotes Radical Prostatectomy.

Cost-Effectiveness

For all screening scenarios, the costs and effects (number of diagnoses, deaths prevented, treatments, life-years and QALYs gained) were compared with a situation without screening. A discount of 3.5% was applied to both costs and effects [15]. Strategies that did not have an alternative or combination of alternatives that would result in more QALYs gained at the same or less net costs were identified as the efficient strategies. For every efficient strategy we determined the incremental cost-effectiveness ratio (ICER), which is calculated as the incremental net costs per incremental QALY gained compared with the previous cost-efficient strategy. The strategy with an ICER value up to a threshold of $100,000 per QALY gained was considered as optimal [16].

Sensitivity Analysis

One-way sensitivity analyses were performed by varying key model parameters. For the utility estimates the highest and lowest values were used (Table 1). A separate analysis was performed using a utility estimate of 1 for the post-recovery period (while retaining all the other utility estimates). The costs were varied by 20%. In addition, the cost-effectiveness was calculated in the absence of overdiagnosis and for a prostate cancer mortality reduction as a result of screening of 56% after 14 years of follow-up, as has been found in the Göteborg trial [17].

Results

Effects of screening ages and interval

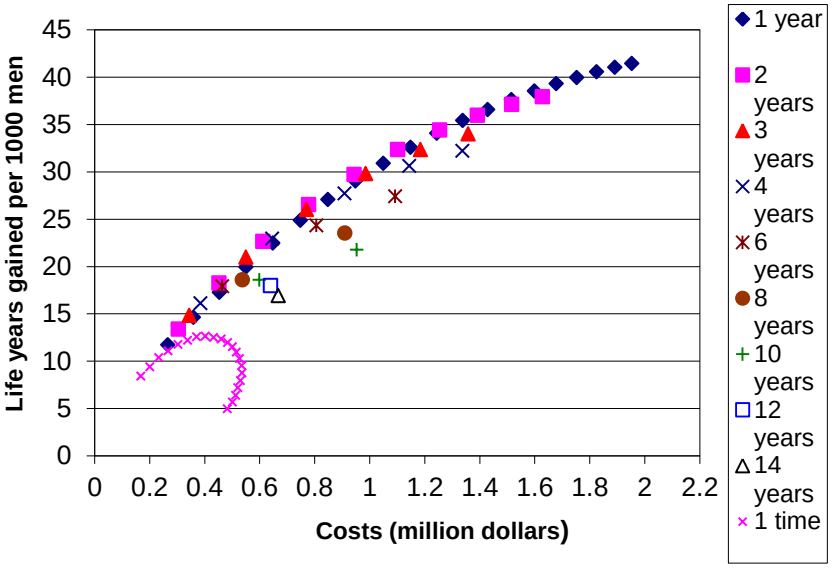
The simulations predicted that without screening 120 per 1000 men would be diagnosed and 32 die from prostate cancer (Table 2). A single screen at age 55 years would result in 4 additional cases diagnosed and 1 prostate cancer death prevented (5% mortality reduction) with 18 life-years gained (17 QALYs) per 1000 men (6.6 quality adjusted days per man).

The cost-effectiveness was $31,467 / QALY gained (3.5% discounted). More intensive screening would increase the number of cancers detected, the mortality reduction, overdiagnosis, the life-years and QALYs gained as well as the costs. The increase in total costs was mainly due to an increase in treatment costs. The largest number of life-years was gained with screening at 1-year intervals at age 55-75 years, but the cost-effectiveness was poor with $320,042 / QALY gained.

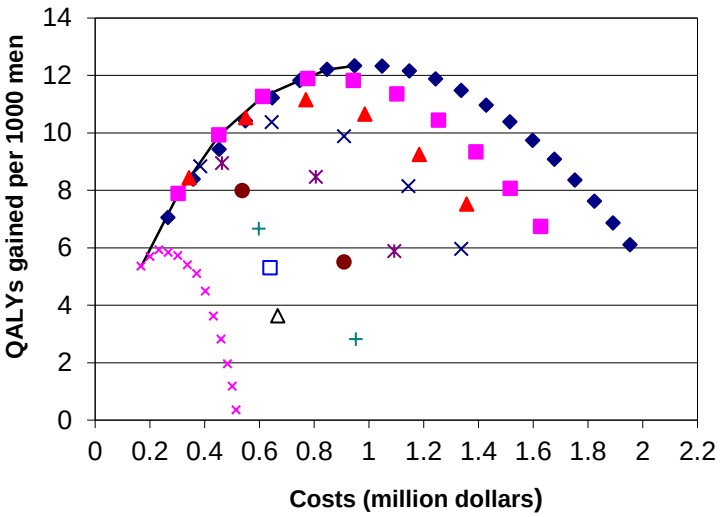
For each level of costs, most life-years were gained with screening at 1- or 2-year intervals (Figure 2A). The largest gain in QALYs was obtained by screening at 1-year intervals from age 55 to 63 years (Figure 1B). For the single screen options, most QALYs were gained by a screen at age 57 years. The strategies on the efficiency frontier (the most effective strategies) had 3, 2 or 1 year intervals. Screening at ages 55-59 years with 2-year intervals yielded an ICER closest to $100,000 per QALY gained and was therefore regarded as optimal (Table 3). Using this strategy of only 3 screens, a 13% prostate cancer mortality reduction was predicted with 33% of the screen-detected cancers overdiagnosed. Using this strategy, the annual death rate is around 25% lower between the ages 60 and 70 when compared with no screening (Figure 3).

Figure 1: Life Years Gained (A) and QALYs Gained (B) against costs for several screening policies.

A



B



*Net costs and (A) life-years gained or (B) quality-adjusted life-years (QALYs) gained (all 3.5% discounted) per 1000 men, of PSA screening strategies varying by interval and end age. The screens start at age 55, except for the once in a life-time screens. At some points in the Figure, the end ages are indicated. The efficient strategies in Figure 1B are connected by the efficient frontier (Eff frontier) and are presented in Table 3. Strategies below this line are less cost-effective.

Table 2. Predicted effects, costs and cost-effectiveness for various screening scenarios per 1000 men. Effects and costs are shown without discount. The cost-effectiveness is calculated at 3.5% discount rate for effects as well as costs.

Screening scenario	No screen	One screen at age 55	Biennial Screening 55-59	Quadrennial Screening 55-67	Biennial Screening 55-69	Screening yearly 55-75
Screening tests	-	800	2,342	2,944	5,706	13,610
Men screened at least once	-	800	935	955	989	997
Effects						
Cancers diagnosed	120	124	132	156	169	207
Screen detected cancers	-	12	34	86	115	180
Overdiagnosed cancers (% of screen-detected)	-	4 (30%)	11 (32%)	35 (41%)	49 (43%)	87 (48%)
Prostate cancer deaths (% reduction)	32	31 (5%)	28 (13%)	25 (24%)	23 (30%)	20 (40%)
Life-years gained	-	18	41	66	83	102
QALYs gained	-	17	36	50	61	64
Costs x $1,000						
Screening	-	32	94	118	228	542
Diagnosis and treatment	1,882	2,003	2,229	2,842	3,161	3,909
Palliative care	649	616	568	496	452	390
Total costs	2,531	2,652	2,890	3,456	3,841	4,842
Cost-effectiveness*						
Net costs per QALY gained (3.5% discounted)	-	31,467	45,615	92,031	120,185	320,042

* The costs and effects are compared with the "no screen" situation, numbers are rounded.

Figure 2: The annual death rate per 1000 men by age in the absence of screening as well as in the presence of screening from age 55 to age 59 with 2 year intervals

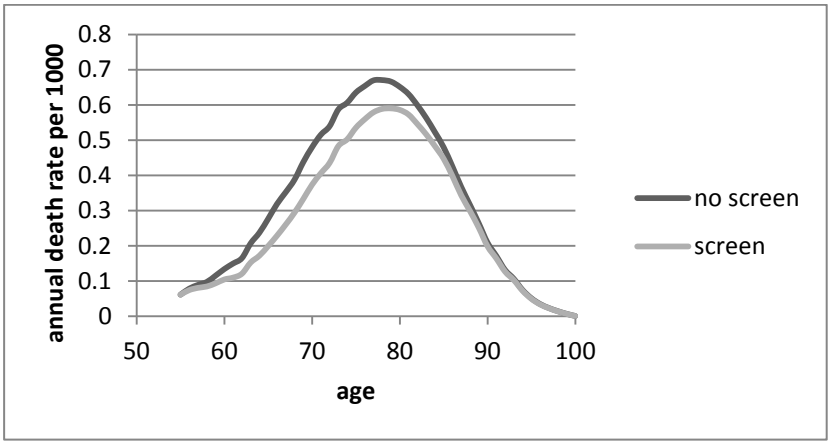


Table 3. Prostate cancer mortality reduction, overdiagnosis, life-years gained and incremental cost-effectiveness for the most efficient screening strategies per 1000 men. The QALYs gained and costs are 3.5% discounted.

Screening strategy		PCM reduction (%)*	LY gained*	Total net costs*	QALY gained*	Cost/ QALY*	ICER*† in $
end age	Freq.						
55	-	5	8.4	168,469	5.4	31,467	31,467
57	2	9	13.4	303,936	7.9	38,563	53,593
58	3	10	14.8	343,908	8.4	40,785	72,567
59	2	13	18.2	452,568	9.9	45,615	72,971
61	2	17	22.6	612,063	11.3	54,349	118,989
61	1	18	24.9	747,784	11.8	63,263	243,031
62	1	20	27.1	848,006	12.2	69,481	260,507
63	1	22	29.0	948,659	12.3	76,910	776,149

* Compared with no screening
† The difference in costs compared with the previous least expensive strategy, divided by the difference in QALYs between those strategies.

Sensitivity Analysis

A sensitivity analysis showed that when all costs for screening, diagnosis and treatment were increased by 20%, the same strategy (age 55-59 years, 2 year interval) remained closest to the optimal ICER (Table 4). Stopping screening at a later age was only favorable when the highest utility estimates were applied, when the utility estimate for the post-recovery period was 1 (no loss in quality of life due to treatment was assumed), when no overdiagnosis would exist or when a mortality reduction of 56% was assumed. In those instances the upper age limit could be 65 to 72 years. When the lowest utility estimates were used, screening at ages 55 and 57 years showed the most favorable cost-effectiveness.

Discussion

Our results suggest that screening strategies with short screening intervals of at most 3 years are more cost-effective than those using longer intervals. Scenarios involving more frequent screening over a limited age range resulted in increased the life-years gained, without a substantial increase in the proportion of overdiagnosed cases. The most favorable results were obtained for screening cessation below age 60 years. The incremental cost-effectiveness ratios of these strategies were $31,467 to $72,971 per QALY gained, close to the commonly used $50,000 and $100,000 thresholds [16].

Table 4. The optimal strategies, with an ICER threshold of $100,000 per QALY gained for various assumptions in the sensitivity analysis. The QALYs gained and costs are 3.5% discounted. Results are presented per 1000 men.

Assumption	Screening strategy		PCM reduction (%)*	Total net costs*	QALY gained *	Cost / QALY *	ICER† in $
	End age	Interval					
Highest utility estimates	72	1	37	1,752,950	51.4	34,122	89,967
Lowest utility estimates	57	2	9	303,936	3.1	98,346	99,913
Utility post-recovery period 1	65	2	24	943,066	25.0	37,664	80,457
Costs – 20%	61	2	17	489,651	11.3	43,479	95,192
Costs + 20%	59	2	13	543,081	9.9	54,739	87,566
No overdiagnosis	71	1	36	542,981	19.4	28,037	93,881
Mortality reduction of 56%	66	1	60	1168,563	46.4	25,205	66,499

* compared with no screening, all costs in US dollars.
† The difference in costs compared with the previous least expensive strategy, divided by the difference in QALYs between those strategies

Earlier we found that men aged 55-59 years with moderate-risk prostate cancer are also the best candidates for immediate curative treatment at the time of screen-detection, because they have the most favorable ratio between lead time and life-years gained [18]. Previous studies concerning the costs or cost-effectiveness of prostate cancer screening have not evaluated life-years gained or QALYs gained [19-24] or were based on assumptions of mortality reduction due to screening, and did not use results of a prostate cancer screening trial to calibrate the model [25-31].

These studies showed large variation in costs-effectiveness from $68 per QALY gained [29] to $729,000 per life-year saved [30], but the results are difficult to compare due to different assumptions in demographics and background risks, screening protocols, costs, effects of treatment and screening on mortality and discount rates. Two studies have used the results of the ERSPC trial to assess cost-effectiveness of screening [6, 32]. They found that screening is not cost-effective with $291,817 per QALY gained and $262,758 per life year gained. Screening can be cost-effective when it is limited to men with 5 times the average risk [6], or when the number needed to treat is less than 18 [32].

Most studies have shown that screening is less cost-effective at higher ages [5]. Our study suggests a lower age at cessation of screening of 59 to 61, whereas previous studies suggest stopping screening at age 70-71 [23, 26, 28, 31]. Our results can change with longer follow-up of the ERSPC trial, as a study in Göteborg suggest that 9 years after termination of screening the prostate cancer mortality in the screen arm caught up [33]. However, the ERSPC has now two additional years of follow-

up, which confirms the relatively stable mortality reduction, compared with the 9 and 11 years follow-up (Schröder et al, in press). The current model can also replicate the mortality reduction after 13-years.

Our conclusion on short intervals may seem surprising. Apparently much of the overdiagnosis is already covered by 4 year interval screening, whereas a shorter interval can still increase the prostate cancer mortality reduction.

Strong points of our study are that the model incorporates a mortality reduction as a result of screening based on a large prostate screening trial and that by simulating a cohort of men for their life-times, all costs and effects can be taken into account.

However, our approach has also some limitations. Since the model is based on the ERSPC trial, in which the majority of men were screened from age 55, the model is not validated to predict results for starting screening below age 55. Several modeling studies have suggested that starting screening at age 40 may improve the cost-effectiveness, or at least lead to comparable prostate cancer mortality reductions with less harms [20, 22, 28]. Also, varying PSA thresholds for biopsy referral for different sub-groups can improve harm-benefit trade-off [20]. For example, higher PSA thresholds can be used for older ages, the screening interval may be based on baseline PSA level, co-morbidity can be taken into account, or other risk stratification methods can be used. We assumed a fixed effect of screening for the entire population. However, this effect can depend on factors such as family history, co-morbidity and ethnicity.

Another limitation of the present study is that most of the disease specific and treatment parameters in the model were based on the data of the ERSPC Rotterdam and the Dutch Cancer Registry, and might not be directly applicable to other populations, especially already more intensively screened populations. Also, the treatments modalities and effects can change in the future. If active surveillance will be used more frequently, the total treatment costs will be lower, whereas an increase in radical prostatectomy or radiotherapy would increase the total costs.

We have not included out-of-pocket costs and indirect costs, such as administrative costs, loss of productivity and income, traveling costs and time and financial losses by family members. Therefore, it is expected that the actual total costs of screening will be higher than predicted in this study. Also in this study cost prices are used whereas reimbursement rates can be higher. Using higher costs would probably not significantly alter the ranking of the results.

The sensitivity analysis showed large differences in cost-effectiveness between the highest and lowest utility estimates. A substantial part of this variation is caused by

the utility estimate for the post-recovery period, because the duration of this health state (the residual life) is around 10 years for most men. Long-term adverse effects from treatment influence the quality of life in the post-recovery period. Quality of life can also be affected more at younger ages than at older ages. However, data for the long-term quality of life after treatment are lacking. Most adverse effects affecting the urinary tract and bowel are relieved after some years, but significant symptoms persist in many patients up to 5 years after treatment [34-37]. In our base model we used a utility estimate of 1 for the time period more than ten years after diagnosis. The use of QALYs to weigh the harms and benefits has been discussed before [4]. Expressing harms and benefits in the same units has been proposed as ideal for providing the evidence base for practice guidelines [38]. When only life years gained are taken into account, the cost-effectiveness is comparable to the cost-effectiveness using the highest utility estimates, and the optimal strategy would be screening from age 55-70 with 1 year intervals.

The AUA recommends shared decision-making for men age 55 to 69 years that are considering PSA screening and does not recommend routine PSA screening in men over age 70 years or less than a 10-15 year life expectancy [39]. AUA also recommend a routine screening interval of two years or more to be preferred over annual screening. Our analysis shows that screening over age 60 years is already less favorable at population level. When screening with 2 year intervals would be stopped at age 59 instead of 69 years, 5 deaths less will be averted but 38 less men will be overdiagnosed, leading to 25 QALYs gained less per 1000 men. Although the AUA and physicians may be reluctant to not recommending screening and shared decision making on the individual level for 60-69 year old men, this analysis provides further evidence of the benefit of going to two year screening intervals.

Our results are more in favor of screening than the report of the USPSTF, recommending against PSA screening [40]. This evaluation was based on a small and inconclusively proven effect of screening, by just summing all prostate cancer screening trials including the PLCO trial, and substantial and well established harms. The PLCO trial had substantial contamination in the control arm [41], negatively effecting the power of the trial, and therefore we based our study on the ERSPC trial.

In conclusion, this analysis based on the largest randomized trial on prostate cancer screening suggests that PSA-based screening can be cost-effective, when it is limited between ages 55 and 60 with intervals of 1 or 2 years. It might be more cost-effective to screen repeatedly between age 55 and 60 with intervals of 1 or 2 years, than using longer intervals until older ages.

References

1. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. N Engl J Med 2012;366(11):981-90.

2. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. J Natl Cancer Inst 2009;101(6):374-83.

3. Heijnsdijk EA, der Kinderen A, Wever EM, et al. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. Br J Cancer 2009;101(11):1833-8.

4. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality-of-life effects of prostate-specific antigen screening. N Engl J Med 2012;367(7):595-605.

5. Garg V, Gu NY, Borrego ME, et al. A literature review of cost-effectiveness analyses of prostate-specific antigen test in prostate cancer screening. Expert Rev Pharmacoecon Outcomes Res 2013;13(3):327-42.

6. Martin AJ, Lord SJ, Verry HE, et al. Risk assessment to guide prostate cancer screening decisions: a cost-effectiveness analysis. Med J Aust 2013;198(10):546-50.

7. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 2003;95(12):868-78.

8. Wever EM, Draisma G, Heijnsdijk EA, et al. How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening. Med Decis Making 2011;31(4):550-8.

9. Wever EM, Hugosson J, Heijnsdijk EA, et al. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. Br J Cancer 2012;107(5):778-84.

10. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. JAMA 2005;293(17):2095-101.

11. Bill-Axelson A, Holmberg L, Filen F, et al. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian prostate cancer group-4 randomized trial. J Natl Cancer Inst 2008;100(16):1144-54.

12. OECD Health Data, Organisation for Economic Co-operation and Development, Pars, France. (Accessed June 5, 2013 at http://www.oecd.org/els/health-systems/OECDHealthData2012FrequentlyRequestedData_Updated_October.xls#'Public).

13. Postma R, Schröder FH, van Leenders GJ, et al. Cancer detection and cancer characteristics in the European Randomized Study of Screening for Prostate Cancer (ERSPC)--Section Rotterdam. A comparison of two rounds of screening. Eur Urol 2007;52(1):89-97.

14. Otto SJ, van der Cruijsen IW, Liem MK, et al. Effective PSA contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. Int J Cancer 2003;105(3):394-9.

15. The guidelines manual January 2009. National Institute for Health and Clinical Excellence, 2009. (Accessed April 28, 2009, at http://www.nice.org.uk/aboutnice/howwework/developingniceclinicalguidelines/clinicalguidelinedevelopmentmethods/GuidelinesManual2009.jsp.).

16. The Cost-Effectiveness Analysis Registry, Center for the Evaluation of Value and Risk in Health. Boston, ICRHPS, Tufts Medical Center. (Accessed June 5, 2013 at http://www.cearegistry.org).

17. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. Lancet Oncol 2010;11(8):725-732.

18. Wever EM, Heijnsdijk EA, Draisma G, et al. Treatment of local-regional prostate cancer detected by PSA screening: benefits and harms according to prognostic factors. Br J Cancer 2013;108(10):1971-7.

19. Bermudez-Tamayo C, Martin Martin JJ, Gonzalez Mdel P, et al. Cost-effectiveness of percent free PSA for prostate cancer detection in men with a total PSA of 4-10 ng/ml. Urol Int 2007;79(4):336-44.

20. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative PSA-based prostate cancer screening strategies. Ann Intern Med 2012;in press.

21. Perez-Niddam K, Thoral F, Charvet-Protat S. Economic evaluation of a prostate cancer screening program in France: a decision model. Crit Rev Oncol Hematol 1999;32(2):167-73.

22. Ross KS, Carter HB, Pearson JD, et al. Comparative efficiency of prostate-specific antigen screening strategies for prostate cancer detection. JAMA 2000;284(11):1399-405.

23. Ross KS, Guess HA, Carter HB. Estimation of treatment benefits when PSA screening for prostate cancer is discontinued at different ages. Urology 2005;66(5):1038-42.

24. Sennfält K, Sandblom G, Carlsson P, et al. Costs and effects of prostate cancer screening in Sweden--a 15-year follow-up of a randomized trial. Scand J Urol Nephrol 2004;38(4):291-8.

25. Barry MJ, Fleming C, Coley CM, et al. Should Medicare provide reimbursement for prostate-specific antigen testing for early detection of prostate cancer? Part IV: Estimating the risks and benefits of an early detection program. Urology 1995;46(4):445-61.

26. Benoit RM, Gronberg H, Naslund MJ. A quantitative analysis of the costs and benefits of prostate cancer screening. Prostate Cancer Prostatic Dis 2001;4(3):138-145.

27. Coley CM, Barry MJ, Fleming C, et al. Early detection of prostate cancer. Part II: Estimating the risks, benefits, and costs. American College of Physicians. Ann Intern Med 1997;126(6):468-79.

28. Howard DH. Life expectancy and the value of early detection. J Health Econ 2005;24(5):891-906.

29. Kobayashi T, Goto R, Ito K, et al. Prostate cancer screening strategies with re-screening interval determined by individual baseline prostate-specific antigen values are cost-effective. Eur J Surg Oncol 2007;33(6):783-9.

30. Krahn MD, Mahoney JE, Eckman MH, et al. Screening for prostate cancer. A decision analytic view. JAMA 1994;272(10):773-80.

31. Zhang J, Denton BT, Balasubramanian H, et al. Optimization of PSA screening policies: a comparison of the patient and societal perspectives. Med Decis Making 2012;32(2):337-49.

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

33. Grenabo Bergdahl A, Holmberg E, Moss S, et al. Incidence of Prostate Cancer After Termination of Screening in a Population-based Randomised Screening Trial. Eur Urol 2013.

34. Johansson E, Steineck G, Holmberg L, et al. Long-term quality-of-life outcomes after radical prostatectomy or watchful waiting: the Scandinavian Prostate Cancer Group-4 randomised trial. Lancet Oncol 2011;12(9):891-9.

35. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med 2008;358(12):1250-61.

36. Smith DP, King MT, Egger S, et al. Quality of life three years after diagnosis of localised prostate cancer: population based cohort study. BMJ 2009;339:b4817.

37. Booth N, Rissanen P, Tammela TL, et al. Health-Related Quality of Life in the Finnish Trial of Screening for Prostate Cancer. Eur Urol 2012.

38. Sox HC. Quality of life and guidelines for PSA screening. N Engl J Med 2012;367(7):669-71.

39. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA guideline. J Urol 2013;13:04308-5.

40. Moyer VA. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2012;157(2):120-134.

41. Gulati R, Tsodikov A, Wever EM, et al. The impact of PLCO control arm contamination on perceived PSA screening efficacy. Cancer Causes Control 2012;23(6):827-35.

Chapter Four

Personalizing Age of Cancer Screening Cessation Based on Comorbidity: Model estimates of harms and benefits

Iris Lansdorp-Vogelaar, Roman Gulati, Angela B. Mariotto, Clyde B. Schechter, Tiago M de Carvalho, Amy B. Knudsen *et al.*

Published: Ann Intern Med. 2014;161(2):104-12

Reproduced with authorization from American College of Physicians

Abstract

Background: Harms and benefits of cancer screening depend on age and comorbidity, yet reliable estimates are lacking.

Objective: To estimate the harms and benefits of cancer screening by age and comorbidity to inform decisions about screening cessation.

Design: Collaborative modeling with seven well-established cancer simulation models and common data on average and comorbidity level-specific life expectancy from SEER-Medicare.

Setting: US population.

Patients: US cohorts aged 66-90 years in 2010 with average health or one of four comorbidity levels (linked to specific conditions): none, mild, moderate, or severe.

Intervention: Mammography, prostate-specific antigen testing, or fecal immunochemical testing.

Measurements: Lifetime cancer deaths prevented and life-years gained (benefits); false-positive tests and overdiagnosed cancers (harms). For each comorbidity level: the age at which harms and benefits of screening were similar to that for individuals with average health undergoing screening at age 74.

Results: Screening 1000 women with average life expectancy at age 74 for breast cancer resulted in 79-96 (range across models) false-positives, 0.5-0.8 overdiagnosed cancers, and 0.7-0.9 breast cancer deaths prevented. While absolute numbers of harms and benefits differed across cancer sites, the ages at which to cease screening were highly consistent across models and cancer sites when based on harm-benefit ratios comparable to screening average-health individuals at age 74. For individuals with no, mild, moderate, and severe comorbidities, screening until ages of

76, 74, 72, and 66, respectively, resulted in similar harms and benefits as for average-health individuals.

Limitations: Comorbidity only influenced life expectancy.

Conclusion: Comorbidity is an important determinant of harms and benefits of screening. Estimates of screening benefits and harms by comorbidity can inform discussions between providers and their older patients about personalizing decisions about when to stop cancer screening.

Introduction

The US Preventive Services Task Force (USPSTF) in 2008 recommended against routine breast and colorectal cancer screening after age 74 (1, 2) because the average gain in life-years associated with extending screening beyond age 74 was felt to be small in comparison to the harms. However, since comorbidity might shift the balance of harms and benefits towards cessation at younger or older ages than 74, the USPSTF and other groups recommended that screening cessation decisions be individualized based on health status (1-4).

There are over 13 million individuals between ages 75 and 85 in the US, and this number is expected to increase to more than 28 million by 2050 (5). Thus, clinicians will be caring for a large and growing number of individuals affected by the uncertainty in how to assess health status and make recommendations regarding screening upper age limits.

There is considerable heterogeneity in the health of these older individuals, yet none of the current guidelines provide clinicians with data to implement personalized approaches. Previous decision analyses looking at health benefits of different cancer screening cessation ages by life expectancy (6, 7) have limited clinical utility because they did not provide a framework for determining life expectancy.

To fill this gap, we estimated the harms and benefits of breast, prostate, and colorectal cancer screening by age based on individual comorbidity using 7 established, independently developed models from the Cancer Intervention and Surveillance Modeling Network (CISNET).

Methods

We used microsimulation models to estimate the harms and benefits of attending one more cancer screen in regularly screened cohorts aged 66-90 by comorbidity level. The harms and benefits for each cohort were compared to that of an average-health cohort attending one more screen at age 74.

The Models

The models used for this analysis are MISCAN-Fadia (MIcrosimulation SCreening Analysis – Fatal diameter) and G-E model (Georgetown-Einstein) for breast cancer; (8, 9) MISCAN-prostate and the FHCRC (Fred Hutchinson Cancer Research Center) model for prostate cancer; (10, 11) and MISCAN-Colon, CRC-SPIN (ColoRectal Cancer Simulated Population model for Incidence and Natural history),

and SimCRC (Simulating ColoRectal Cancer) for colorectal cancer (12-14).

Each model simulates the life histories of individuals from birth to death and tracks underlying disease in the presence and absence of screening. These models have previously been applied to inform the USPSTF recommendations for breast and colorectal cancer screening (15, 16) and to evaluate prostate cancer screening and treatment interventions (17, 18). Using multiple models per cancer site provides a credible range of results and serves as a sensitivity analysis on the impact of variations in underlying model structure and assumptions (Supplementary Information Table 1).

Briefly, screening extends life through detection of disease at an earlier stage or a smaller size when it may have better survival after treatment than in the absence of screening. Inputs were standardized across models within cancer site, including test characteristics, screening and follow-up assumptions, treatment distributions, and cancer-specific and other-cause survival. Sources for the model inputs have been described in prior publications (15, 16, 19).

Descriptions of each model have been published elsewhere (8-14); model profiles are available (http://cisnet.cancer.gov/profiles/) and additional information about the models is available from the authors upon request. CISNET also includes procedures for external collaboration for interested investigators (http://cisnet.cancer.gov). Selected model outputs for 74-year old individuals who have average health (and life expectancy) and who had been screened regularly prior to age 74 are provided in Table 1.

Population

After assuming that all individuals underwent regular screening starting at age 50 with biennial mammography, biennial prostate-specific antigen (PSA) testing, or annual fecal immunochemical testing (FIT), respectively, the models begin their simulation for US cohorts of individuals aged 66 to 90 years old in the year 2010 who have average risk of cancer and a specified comorbidity level (none, mild, moderate, or severe) and follow these individuals for their remaining lifetime. For reference, we also simulated cohorts aged 74 years and 76 years (75 years for colorectal cancer) with average health and corresponding average life expectancy. We assumed that comorbidity influenced non-cancer life expectancy but not cancer risk or progression, treatment, or cancer-specific survival.

Table 1. Selected clinical model outputs for 74-year old individuals who have average health (and life expectancy) and who had been screened regularly prior to age 74, in the presence and absence of screening at age 74.

	Breast Cancer		Prostate Cancer		Colorectal Cancer		
	MISCAN-Fadia	**G-E model**	**MISCAN-Prostate**	**FHCRC Prostate**	**MISCAN-Colon**	**CRC-SPIN**	**SimCRC**
Lifetime probability of developing cancer without screening at age 74	6.9%	8.2%	7.9%	4.7%	1.9%	1.3%	1.1%
Lifetime probability of developing cancer with screening at age 74	7.2%	8.4%	9.8%	6.1%	1.8%	1.2%	1.0%
Prevalence of undiagnosed cancer immediately before a screening at age 74	9.3%*	1.14%	4.4%	13.3%	0.19%	0.06%	0.09%
Prevalence of undiagnosed cancer immediately after a screening at age 74	8.2%*	0.09%	1.0%	11.3%	0.06%	0.02%	0.05%

* The large difference in undiagnosed cancers between the models reflect differences in modeling approach. MISCAN-Fadia includes cancers from a very small tumor size (0.1 mm) not yet detectable by mammography.
MISCAN: Microsimulation Screening Analysis; Fadia: Fatal diameter; G-E: Georgetown-Einstein; FHCRC: Fred Hutchinson Cancer Research Center; CRC-SPIN:ColoRectal Cancer Simulated Population model for Incidence and Natural history; SimCRC: Simulating ColoRectal Cancer

Comorbidity-specific life tables

Non-cancer life expectancy was derived from comorbidity scores for 16 conditions derived from claims from a random 5% sample of non-cancer beneficiaries continuously enrolled with Medicare Parts A and B from 1992 to 2005 and residing in the Surveillance Epidemiology, and End Results (SEER) areas (20, 21, 22). Cox proportional hazard methods were used to estimate non-cancer age-conditional life tables for each gender and age combination using comorbidity as a covariate.

Comorbidity was then grouped into four levels: none, mild, moderate, and severe, each with its own life expectancy at a given age (Table 2). We used the weighted average of the comorbidity-specific life tables for the reference average-health cohorts. We extrapolated beyond the 13 years of available data by assuming that

mortality rates converged abruptly to average US rates after that period. This is a conservative assumption for personalizing age of screening cessation because life expectancy is overestimated for those with moderate and severe comorbidity and is underestimated for those with no or mild comorbidities.

Table 2 Overview of comorbidity levels, associated conditions, and life expectancies at ages 68, 74 and 78 years

Comorbid group	% of population at age 74	Conditions included*	HR†	Life expectancy at age 68		Life expectancy at age 74		Life expectancy at age 78	
				M	W	M	W	M	W
No	69%	None	1	16.6	19.3	13.1	15.1	10.7	12.4
Mild	2%	History of MI, acute MI, ulcer or rheumatologic disease	1.01 - 1.38	15.4	17.1	12.5	13.1	9.8	10.9
Moderate	12%	(Cardio-)vascular disease; paralysis; diabetes; or combinations of diabetes with MI, ulcer, or rheumatologic disease	1.39 - 1.66	14.4	16.3	11.0	12.4	8.6	9.9
Severe	17%	AIDS; COPD; mild or severe liver disease; chronic renal failure; dementia; congestive heart failure; or combinations of aforementioned diseases not categorized under moderate comorbidity	≥ 1.67	10.8	13.3	8.1	9.8	6.4	7.7
Average health	100%	All				11.9	13.9		

HR: Hazard ratio; y year; MI myocardial infarction; AIDS Acquired Immune Deficiency Syndrome; COPD chronic obstructive pulmonary disease

* Any one of the conditions listed places an individual in the associated comorbidity level in the first column of the table and its life expectancy at age 68, 74 or 78. Conditions included are those that affect life expectancy. See Appendix Table 2 for life expectancies at other ages. See Appendix Table 3 for ICD-09 codes for the conditions.

† Hazard ratio for all-cause mortality compared to no comorbidity for the conditions included in each comorbidity level

Analysis

For each cohort, we estimated the benefits and harms of screening at their current age. Diagnostic follow-up was based on current recommendations (breast and colorectal) or practice (prostate). Screening benefits were expressed as the life-years gained (LYG) and cancer deaths prevented (CDP) for every 1000 individuals screened at a given age. Harms were expressed as the false-positive tests and overdiagnosed cancers (i.e., cancer that would not have caused symptoms during an individual's lifetime) per 1000 individuals screened.

The balance between harms and benefits was expressed as the number needed to screen to gain one life-year (NNS/LYG). For reference, we also determined, for each comorbidity level, the age at which the harms and benefits of screening (NNS/LYG) were similar to screening the average-health population at age 74.

Sensitivity Analysis

We varied our method for extrapolating comorbidity-specific life tables by assuming that the hazard ratio between the average-health life table and the comorbidity-specific life table at the 13th year of observation was maintained until death. For colorectal cancer, we also estimated the harms and benefits of colonoscopy screening by age and comorbidity level.

Results

Screening Based on "Average" Comorbidity Level

At age 74, the average life expectancy for women across all comorbidity groups is 13.9 years. Screening 1000 women for breast cancer who have average health (and life expectancy) and who had been screened regularly prior to age 74 resulted in 79-96 false-positive tests (range across models) and 0.5-0.8 overdiagnosed cancers (Table 3). On the benefits side, 0.7-0.9 cancer deaths would be prevented and 5.8-7.6 life years gained (LYG), corresponding to 132-173 women that need to be screened at age 74 to gain one life-year.

Screening 1000 women age 76 and older for breast cancer yields increased harms and decreased benefits. Specifically, 146-198 women needed to be screened at age 76 to gain one life-year. The balance of harms and benefits for prostate and colorectal cancer screening were mostly comparable except for the rates of overdiagnosis, which were orders of magnitude (15-100+ times, depending on the model) higher for prostate cancer vs. breast or colorectal cancer screening.

Table 3. Benefits and harms of screening 1,000 regularly screened individuals at age 74 or 76 (75 years for colorectal cancer) with average health, by cancer site and model

Cancer site / model / age of screening	Harms*		Incremental benefits*		Balance	
	False-positive tests	Over-diagnosed cancers	Life-years gained†	Cancer deaths prevented	NNS/LYG	NNS/CDP
Breast cancer						
MISCAN-Fadia						
74 years	79	0.8	7.6	0.9	132	1125
76 years	77	1.0	6.9	0.9	146	1102
G-E model						
74 years	96	0.5	5.8	0.7	173	1421
76 years	96	0.6	5.1	0.7	198	1474
Prostate cancer						
MISCAN-prostate						
74 years	116	19.7	6.6	1.2	150	830
76 years	136	24.5	6.3	1.2	159	820
FHCRC prostate cancer model						
74 years	242	14.5	6.1	0.8	165	1263
76 years	268	16.2	5.1	0.7	197	1371
Colorectal cancer						
MISCAN-Colon						
74 years	39	0.3	6.2	0.9	161	1118
75 years	39	0.4	5.5	0.8	182	1218
CRC-SPIN						
74 years	38	0.0	3.9	0.7	256	1518
75 years	38	0.0	3.9	0.6	254	1629
SimCRC						
74 years	38	0.1	4.9	0.8	227	1411
75 years	38	0.1	4.3	0.7	258	1522

* Results are per 1,000 individuals screened according to guidelines (breast and colorectal) or current practice (prostate) since age 50

† One life-year gained per 1,000 individuals corresponds with 0.4 days gained per individual. NNS/LYG: Number needed to screen to gain 1 life-year; NNS/CDP: Number needed to screen to prevent 1 cancer death; MISCAN: Microsimulation Screening Analysis; Fadia: Fatal diameter; G-E: Georgetwon-Einstein; FHCRC: Fred Hutchinson Cancer Research Center; CRC-SPIN:ColoRectal Cancer Simulated Population model for Incidence and Natural history; SimCRC: Simulating ColoRectal Cancer

Screening by Comorbidity Level

In individuals with no comorbidities (i.e., individuals with longer-than-average life expectancy), screening 1000 regularly screened women aged 74 resulted in fewer overdiagnosed breast cancers (0.3-0.5) and more cancer deaths prevented (0.8-1.0) and LYG (6.6-8.5) compared to women the same age with average health.

Consequently, the NNS/LYG of 117–150 was lower than that for the average-health population (Appendix Table 2). In fact, women with no comorbidities could be screened until age 76-78 and still yield a similar NNS/LYG as screening until age 74 in the average-health population (Figures 1 and 2). The same result is obtained for prostate and colorectal cancer screening (Figure 1, Appendix Table 2). For the mild comorbidity group, screening for all three cancers at age 74 yielded similar harms and benefits as screening the average-health population (Figures 1 and 2, Appendix Table 2).

In individuals with moderate comorbidity, screening at age 74 was considerably less favorable than in the average-health population at that age: overdiagnosed cancers were up to 15% higher, while cancer deaths prevented and LYG were up to 20% lower (Appendix Table 2). Screening those with moderate comorbidity at a median age of 72 (range 68-74) resulted in similar harms and benefits as screening the average-health population at age 74 (Figures 1 and 2, Appendix Table 2). Screening people with severe comorbidities for breast cancer at age 74 resulted in even more harms (1.3-1.9 overdiagnosed cancers) and fewer benefits (0.5-0.6 cancer deaths prevented; 3.5-4.5 LYG) (Appendix Table 2). In this group, screening at a median age of 66 (range 64–69) provided similar harms and benefits as screening the average-health population at age 74 (Figures 1 and 2, Appendix Table 2).

Sensitivity Analysis

If non-cancer mortality rates do not converge to average US rates after the 13 years of observed data, individuals with moderate and severe comorbidity could stop screening at even younger ages to have similar harms and benefits as screening the average-health population at age 74. `

For colonoscopy, screening to age 70 for those with no, mild, or moderate comorbidity provided similar harms and benefits as screening the average-health population at age 70 (the last colonoscopy screening age for an individual regularly screened since age 50); among those with severe comorbidity the balance is comparable at age 60 (Appendix Table 4).

Discussion

This is the first study to employ collaborative modeling to evaluate screening across three cancer sites. It systematically quantifies the balance of benefits and harms of screening older individuals for breast, prostate, and colorectal cancers by comorbidity level. The results are robust across models and cancer sites and indicate that comorbidity affects screening benefits and harms and decisions about ages of

screening cessation. These outcomes can directly inform individualized decisions about undergoing screening. Around 70% of the current US population aged 74 have none of the comorbidities noted in recent analyses to influence life expectancy (20). Our results suggest that this group could continue to be screened until age 76 and still have the same balance of benefits and harms expected from screening the average-health population until age 74. However, the 13% of the US population aged 65 to 74 with severe comorbidity should stop screening at age 66 to have the same balance of benefits and harms as seen among average-health groups undergoing screening from ages 50 to 74.

Our findings are consistent with and extend prior research addressing the upper age limits for cancer screening. For instance, Walter and Covinsky (7) found that screening for breast, cervical, and colorectal cancer until around age 60 for individuals in the lower quartile of population life expectancy has the same number needed to screen to prevent one cancer death as those with the median life expectancy at age 75. They also estimated that screening could be continued to age 85 for those in the upper quartile of life expectancy (7). This range is consistent with but wider than our model projections of 66 to 76 based on number needed to screen per life-year gained. When we use cancer deaths prevented rather than LYG, our ranges were closer to those of Walter and Covinsky, although our maximal upper age limit was still lower since we considered individuals without comorbidities (70% of the population) and they used the upper quartile (25%) of life expectancy.

A recent analysis looking at time lag to benefit after screening for breast and colorectal cancer suggests that screening for breast and colorectal cancer is most appropriate for patients with a life expectancy greater than 10 years (23). However, that study and others (6, 7, 24-29) provide little guidance on applying this framework in clinical practice, leaving it to clinical judgment to estimate life expectancy and individualize screening decisions. Several studies have investigated the relationship between comorbidity level and life expectancy (30-32) but do not address the question of how this relationship influences cancer screening. To date, only two analyses have directly related comorbidity level to cancer screening recommendations. One focused only on diabetes-related comorbidities and colorectal cancer screening (33) and one examined cardiovascular disease and breast screening (6). The current analysis is a multi-model collaborative analysis of three major cancers and includes a wider range of comorbidities than considered previously.

There is considerable debate about the value of PSA screening. The USPSTF recently concluded that PSA screening results in little or no reduction in prostate cancer-specific mortality while leading to substantial prostate cancer overdiagnosis.

Based on these findings, the USPSTF recommends against routine prostate cancer screening. However, like other national guidelines panels (e.g., the ACS, the American Urological Association, the American College of Physicians), they recognized a role for screening in the context of appropriate patient-physician decision-making (34-37). Our modeling study provides clinicians with valuable information for such shared-decision making. Recent modelling studies indicate that overdiagnoses and unnecessary biopsies could be reduced by stopping prostate cancer screening at age 69, raising the PSA threshold for biopsy referral for men above this age, or restricting further screening to men with low comorbidity (17, 18). Our results confirm these results and underscore that overdiagnosed cancers detected via PSA screening are orders of magnitude higher than for breast or colorectal cancer screening. If personalized PSA screening strategies achieve a sufficiently favorable balance of outcomes, our results advocate tailoring of screening cessation according to comorbidity-based life expectancy.

Our results provide clinicians with data for use in informed decision-making discussions about who might consider continuing screening and for how long. For example, if a clinician is meeting with a regularly-screened 70-year old patient with COPD, our results indicate that this individual falls in the severe comorbidity level and, depending on patient preferences, the benefits of screening may no longer outweigh the potential harms. However, a 76-year old individual with no comorbidities might consider attending another screen. The final decision about screening should depend on individual patient preferences. Prevention of death from cancer is one outcome to consider, but some patients may be more concerned with impact on other outcomes such as quality of life or functional independence. Individuals may also prefer to be detected at earlier stages when less intensive treatment may be needed, compared to somewhat later diagnosis and more aggressive therapy, even if survival is unchanged. In such a situation, a healthy patient may choose to continue (for example) breast cancer screening up to age 80, where the mortality benefit is small, but early detection can find the cancer early when less aggressive treatment is required.

The fact that comorbidity-specific conclusions about age-specific benefits and harms differ meaningfully from those included in clinical guidelines highlights the tension between the need to provide public health recommendations for the general population and the potential advantages of using a more personalized approach. Our suggested approach of continuing screening in the healthy and earlier cessation in the sickest individuals does not increase the number of screens required in the population but rather leads to a more efficient allocation of resources, increasing the benefit and decreasing the harms to the growing older population (38). The age-, gender-, and comorbidity-specific life expectancies used for these analyses also provide clinicians and the general screening-eligible population with a foundation

for discussing preferences for benefits and harms, facilitating individual decision-making.

The testing intervals for colorectal and breast cancer were chosen according to the latest USPSTF guidelines, but our conclusions can be generalized to other intervals. The screening cessation ages are determined in relation to screening the average-health population up to age 74. When choosing a different screening interval, such as annual mammography for breast cancer, the benefits and harms will be different for screening the average-health population at age 74. The benefits and harms by comorbidity level will change accordingly, such that the optimal screening cessation age by comorbidity remains the same.

Despite the innovation and strengths of our approach, there are several caveats that should be considered in evaluating our results. First, we chose the balance of harms and benefits as our primary metric. We did not explicitly consider complications from screening and diagnostic follow-up as harms, but these would be proportional to the number of (false-positive) screening tests. Costs per quality-adjusted life-year gained are another common metric used in many countries, but this is not widely accepted in the US (39-43). Second, we assumed that comorbidity level only influenced life expectancy and not cancer risk or biology. Health conditions such as diabetes are known to be associated with obesity and other lifestyle factors (44) which, in turn, can be associated with improved mammography performance (45) and elevated breast (46-48) or reduced prostate (49) cancer risk. Conversely, adverse events of screening, such as perforations with colonoscopy, are also associated with comorbidities (50).

In the future, it will be important to extend our work to capture the known impact of specific comorbidities on other model parameters. For now, competing non-cancer mortality is the single most germane parameter in screening decisions for the oldest age groups so that our conclusions should be robust (27). Third, we only estimated harms and benefits of screening by comorbidity for people aged 66 years and older because life expectancy estimates by comorbidity were obtained from SEER-Medicare data.

Next, we only considered individuals regularly screened since age 50 to demonstrate how current screening recommendations could be adapted based on comorbidity level. In general, stopping ages are higher in individuals who are unscreened or have skipped previous screening rounds because they have a higher risk of prevalent cancer. Furthermore, the models used life tables based on non-cancer cases and, therefore, do not include cancer-specific mortality for cancers other than the one targeted by screening. This underestimates the true rate of competing other-cause mortality, and therefore the harm-benefit ratios, but does not

affect our internal comparisons of comorbidity level groups to the average-health population. Finally, we did not consider situations in which the comorbidity level decreased (e.g., from severe to moderate level) after the age of screening cessation. Given the chronic nature of the comorbid conditions in older ages, this is a reasonable assumption.

Overall, the results across models and cancer sites were very robust and strongly suggest that the age of screening cessation based on comorbidity levels varies by nearly a 10-year interval around the age cut-point of 74 included in current breast and colorectal cancer screening recommendations. Our data on common chronic health conditions and their associated comorbidity level, together with model projections of screening benefits and harms at each of these comorbidity levels, can inform discussions between providers and their older patients about personalizing decisions about when to stop cancer screening.

References

1. U. S. Preventive Services Task Force. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2008 Nov 4;149(9):627-37.

2. U. S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009 Nov 17;151(10):716-26, W-236.

3. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. J Urol. 2013 Aug;190(2):419-26.

4. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2012: A review of current American Cancer Society guidelines and current issues in cancer screening. CA Cancer J Clin. 2012;62:129-42.

5. 2012 National Population Projections: Table 1. Projected Population by Single Year of Age, Sex, Race, and Hispanic Origin for the United States: 2012 to 2060 [database on the Internet]. U.S. Census Bureau, Population Division. 2012. Accessed at http://www.census.gov/population/projections/data/national/2012/downloadablefiles.html on March 17, 2013.

6. Mandelblatt JS, Wheat ME, Monane M, et al. Breast cancer screening for elderly women with and without comorbid conditions. A decision analysis model. Ann Intern Med. 1992 ;116(9):722-30.

7. Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA. 2001;285(21):2750-6.

8. Tan SY, van Oortmarssen GJ, de Koning HJ, et al. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr. 2006(36):56-65.

9. Mandelblatt J, Schechter CB, Lawrence W, et al. The SPECTRUM population model of the impact of screening and treatment on U.S. breast cancer trends from 1975 to 2000: principles and practice of the model methods. J Natl Cancer Inst Monogr. 2006(36):47-55.

10. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003;95(12):868-78.

11. Gulati R, Inoue L, Katcher J, et al. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics. 2010;11(4):707-19.

12. Frazier AL, Colditz GA, Fuchs CS, et al. Cost-effectiveness of screening for colorectal cancer in the general population. JAMA. 2000;284(15):1954-61.

13. Rutter CM, Savarino JE. An evidence-based microsimulation model for colorectal cancer: validation and application. Cancer Epidemiol Biomarkers Prev. 2010;19(8):1992-2002.

14. Loeve F, Boer R, van Oortmarssen GJ, et al. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. Comput Biomed Res. 1999 Feb;32(1):13-33.

15. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under

different screening schedules: model estimates of potential benefits and harms. Ann Intern Med. 2009 ;151(10):738-47.

16. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, et al. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med. 2008 ;149(9):659-69.

17. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: Model estimates of potential benefits and harms. Ann Intern Med. 2013;173(3):227-8.

18. Heijnsdijk EAM, Wever EM, Auvinen A, et al . Quality-of-life effects of prostate-specific antigen screening. New Engl J Med. 2012;367:595-605.

19. Gulati R, Wever EM, Tsodikov A, et al. What if i don't treat my PSA-detected prostate cancer? Answers from three natural history models. Cancer Epidemiol Biomarkers Prev. 2011;20(5):740-50.

20. Cho H, Klabunde CN, Yabroff KR, , et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. Ann Intern Med. 2013;159(10):667-76.

21. Klabunde CN, Legler JM, Warren JL, et al. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. Ann Epidemiol. 2007;17(8):584-90.

22. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. J Clin Epidemiol. 2000;53(12):1258-67.

23. Lee SJ, Boscardin WJ, Stijacic-Cenzer I, et al. Time lag to benefit after screening for breast and colorectal cancer: meta-analysis of survival data from the United States, Sweden, United Kingdom, and Denmark. BMJ. 2012;345(doi: 10.1136/bmj.e8441):1-9.

24. Harewood GC, Lawlor GO, Larson MV. Incident rates of colonic neoplasia in older patients: when should we stop screening? J Gastroenterol Hepatol. 2006;21(6):1021-5.

25. Ko CW, Sonnenberg A. Comparing risks and benefits of colorectal cancer screening in elderly patients. Gastroenterology. 2005;129(4):1163-70.

26. Maheshwari S, Patel T, Patel P. Screening for colorectal cancer in elderly persons: who should we screen and when can we stop? J Aging Health. 2008;20(1):126-39.

27. Mandelblatt JS, Schechter CB, Yabroff KR, et al. Toward optimal screening strategies for older women. Costs, benefits, and harms of breast cancer screening by age, biology, and health status. J Gen Intern Med. 2005;20(6):487-96.

28. Rich JS, Black WC. When should we stop screening? Eff Clin Pract. 2000;3(2):78-84.

29. Stevens T, Burke CA. Colonoscopy screening in the elderly: when to stop? Am J Gastroenterol. 2003;98(8):1881-5.

30. Schonberg MA, Davis RB, McCarthy EP, et al. External validation of an index to predict up to 9-year mortality of community-dwelling adults aged 65 and older. J Am Geriatr Soc. 2011 ;59(8):1444-51.

31. Schonberg MA, Davis RB, McCarthy EP, et al. Index to predict 5-year mortality of community-dwelling adults aged 65 and older using data from the National Health Interview Survey. J Gen Intern Med. 2009;24(10):1115-22.

32. Yourman LC, Lee SJ, Schonberg MA, et al. Prognostic indices for older adults: a systematic review. JAMA. 2012;307(2):182-92.

33. Dinh TA, Alperin P, Walter LC, et al. Impact of comorbidity on colorectal cancer screening cost-effectiveness study in diabetic populations. J Gen Intern Med. 2012;27(6):730-8.

34. Greene KL, Albertsen PC, Babaian RJ, et al. Prostate specific antigen best practice statement: 2009 update. J Urol. 2009 ;182(5):2232-41.

35. Kawachi MH, Bahnson RR, Barry M, et al. Prostate cancer early detection. Clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2007;5(7):714-36.

36. Qaseem A, Barry MJ, Denberg TD, et al, Clinical Guidelines Committee of the American College of P. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. Ann Intern Med. 2013;158(10):761-9.

37. Wolf AM, Wender RC, Etzioni RB, , et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J Clin. 2010;60(2):70-98.

38. Mandelblatt J, Tosteston AN, van Ravesteyn NT. Costs, evidence and value in the Medicare program: the challenges of technology innovation in breast cancer prevention and control. JAMA Intern Med. 2013;173(3):227-8.

39. Begg C. Comments and response on the USPSTF recommendation on screening for breast cancer. Ann Intern Med. 2010;152(8):540-1; author reply 3-4.

40. Braithwaite R. Comments and response on the USPSTF recommendation on screening for breast cancer. Ann Intern Med. 2010;152(8):539-40; author reply 43-4.

41. Col N, Hansen MH, Fischhoff B, et al. Comments and response on the USPSTF recommendation on screening for breast cancer. Ann Intern Med. 2010;152(8):542; author reply 3-4.

42. Harris R, Sawaya GF, Moyer VA, et al. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive Services Task Force. Epidemiol Rev. 2011;33(1):20-35.

43. Ho A. Comments and response on the USPSTF recommendation on screening for breast cancer. Ann Intern Med. 2010;152(8):542-3; author reply 3-4.

44. Klijs B, Mackenbach JP, Kunst AE. Obesity, smoking, alcohol consumption and years lived with disability: a Sullivan life table approach. BMC Public Health. 2011;11:378.

45. Chang Y, Schechter C, van Ravesteyn N, et al J. Collaborative Modeling of the Impact of Obesity on Race–specific Breast Cancer Incidence and Mortality. Breast Cancer Research and Treatment. 2012;DOI: 10.1007/s10549-012-2274-3.

46. Huxley RR, Ansary-Moghaddam A, Clifton P, et al. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer. 2009;125(1):171-80.

47. Nelson NJ. Studies on how lifestyle factors may affect breast cancer risk and recurrence. J Natl Cancer Inst. 2012;104(8):574-6.

48. Wolk A. Diet, lifestyle and risk of prostate cancer. Acta Oncol. 2005;44(3):277-81.

49. Gong Z, Neuhouser ML, Goodman PJ, et al. Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. Cancer Epidemiol Biomarkers Prev. 2006;15(10):1977-83.

50. Warren JL, Klabunde CN, Mariotto AB, , et al. Adverse events after outpatient colonoscopy in the Medicare population. Ann Intern Med. 2009;150(12):849-57, W152.

Supplementary Information

Table 1. Summary of model features by cancer site and model

	Breast Cancer		Prostate Cancer		Colorectal Cancer		
	MISCAN-Fadia	**G-E model**	**MISCAN-Prostate**	**FHCRC Prostate**	**MISCAN-Colon**	**CRC-SPIN**	**SimCRC**
Modeling software	Borland Delphi	C++	Borland Delphi	C	Borland Delphi	Microsoft C#	Microsoft Visual C++ 2010 Express
Includes pre-cancers *	Yes	Yes	No	No	Yes	Yes	Yes
Includes tumor biomarkers †	Yes	Yes	Yes	Yes	No	No	No
Calibrated to incidence?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Calibrated to mortality? §	No	No	No	No	No	No	No
How treatment affects mortality	Cure fraction	Hazard reduction	Hazard reduction	Hazard reduction	Not explicitly modeled ‡	Not explicitly modeled ‡	Not explicitly modeled ‡
How screening affects mortality	Cancer diameter shift, age-shift	Stage-shift, age-shift	Cure fraction	Stage-shift, age-shift	Stage-shift, age-shift	Stage-shift, age-shift	Stage-shift, age-shift

* Pre-cancers include ductal carcinomas in situ for breast cancer and adenomas for colorectal cancer.
† Tumor biomarkers include estrogen receptor / human epidermal growth factor 2 status for breast cancer, and prostate-specific antigen level and Gleason score for prostate cancer
‡ The models use the latest relative survival estimates from SEER to model the probability of dying from colorectal cancer
§ The models use incidence and relative survival estimates to match observed mortality

Table 2a. Benefits and harms of screening 1,000 regularly screened individuals by age and comorbidity – MISCAN-prostate for prostate cancer.

Comorbidity / age of screening	Life-expectancy (years)	Harms*		Incremental benefits*		Balance	
		False-positive tests	Over-diagnosed cancers	Life-years gained†	Cancer deaths prevented†	NNS/LYG†	NNS/CDP†
No comorbidity							
66	17.8	69.2	8.6	8.7	1.1	115	885
68	16.6	76.0	10.0	8.3	1.2	120	843
70	15.5	82.7	11.5	7.9	1.2	127	831
72	14.4	95.3	14.2	7.5	1.3	133	796
74	13.1	116.3	18.4	7.8	1.4	129	722
76	11.9	135.7	22.9	7.4	1.4	135	704
78	10.7	149.3	26.8	6.2	1.3	162	770
80	9.6	157.5	29.9	5.0	1.1	199	877
82	8.5	160.1	32.1	3.6	0.9	276	1073
84	7.5	156.6	33.0	2.4	0.7	410	1463
86	6.5	147.2	32.5	1.5	0.5	690	2116
88	5.8	134.4	31.0	0.9	0.3	1143	3073
90	5.1	118.9	28.3	0.5	0.2	1957	4718
Mild							
66	16.7	69.2	9.2	7.9	1.0	126	967
68	15.4	76.0	10.8	7.2	1.0	138	959
70	14.6	82.6	12.1	7.1	1.1	140	904
72	13.4	95.3	15.0	6.4	1.1	157	907
74	12.5	116.3	19.0	7.0	1.3	142	775
76	11.7	135.7	23.3	7.4	1.4	136	716
78	9.8	149.3	28.0	5.0	1.1	198	900
80	9.1	157.5	30.7	4.4	1.0	229	971
82	8.3	160.1	32.5	3.1	0.9	320	1165
84	7.3	156.6	33.4	2.4	0.7	422	1500
86	6.2	147.2	33.0	1.3	0.4	768	2305
88	5.7	134.4	30.9	0.8	0.3	1179	3042
90	4.9	118.9	28.3	0.5	0.2	2091	4876

Table 2b. Benefits and harms of screening 1,000 regularly screened individuals by age and comorbidity – MISCAN-prostate for prostate cancer.

Moderate comorbidity§							
66	15.3	69.2	10.0	6.8	0.9	147	1108
68	14.4	76.0	11.5	6.4	0.9	156	1072
70	13.0	82.6	13.3	5.6	0.9	178	1113
72	11.8	95.3	16.4	5.0	0.9	200	1109
74	11.0	116.3	20.7	5.4	1.0	184	978
76	9.7	135.7	25.7	4.8	1.0	207	1012
78	8.6	149.3	29.7	3.8	0.9	266	1135
80	7.7	157.5	32.9	2.9	0.8	342	1318
82	6.7	160.0	35.1	2.0	0.6	495	1678
84	6.0	156.6	35.6	1.4	0.4	709	2261
86	5.4	147.2	34.3	1.0	0.3	1038	2929
88	4.7	134.4	32.3	0.6	0.2	1734	4225
90	3.9	119.0	29.8	0.3	0.1	3302	7065
Severe comorbidity‖							
66	11.6	69.2	12.3	4.5	0.6	220	1621
68	10.8	76.0	14.0	4.2	0.6	241	1598
70	9.8	82.6	15.8	3.5	0.6	282	1688
72	9.0	95.3	19.0	3.2	0.6	310	1652
74	8.1	116.3	23.9	3.3	0.6	301	1551
76	7.2	135.7	29.2	2.9	0.6	349	1593
78	6.4	149.3	33.2	2.3	0.6	435	1778
80	5.8	157.5	36.2	1.8	0.5	550	2028
82	5.1	160.1	37.9	1.2	0.4	815	2658
84	4.4	156.5	38.2	0.8	0.3	1208	3572
86	4.0	147.2	36.6	0.6	0.2	1801	4770
88	3.4	134.4	34.3	0.3	0.1	3321	7404
90	3.1	119.0	30.7	0.2	0.1	5017	10382

NNS/LYG: Number needed to screen to gain 1 life-year; NNS/CDP: Number needed to screen to prevent 1 cancer death
Shaded row represents the age for each comorbidity group at which screening provided similar harms and benefits as screening at age 74 in the entire population

* Results are per 1,000 individuals screened according to current practice since age 50
† Irregular pattern by age is caused by stepwise increase in onset of prostate cancer by age

‡ Mild comorbidity includes having a life expectancy associated with having a history of MI, acute MI, ulcer, or rheumatologic disease
§ Moderate comorbidity includes having a life-expectancy associated with having vascular disease, cardiovascular disease, paralysis or diabetes, or combinations of diabetes with MI, ulcer, or rheumatologic disease
|| Severe comorbidity includes having a life expectancy associated with having AIDS, COPD, mild liver disease, severe liver disease, chronic renal failure, dementia, or congestive heart failure, or combinations of aforementioned diseases not categorized under moderate comorbidity

Chapter Five

Is prostate cancer different in black men? Answers from three natural history models

Alex Tsodikov, Roman Gulati, Tiago M. de Carvalho, Eveline A. M. Heijnsdijk, Rachel A. Hunter-Merrill, Angela B. Mariotto, *et al*.

Accepted for publication in Cancer

Reproduced with authorization from Wiley-Blackwell

Abstract

Background: Black men in the US have substantially higher prostate cancer incidence rates than the general population. The extent to which the incidence disparity is due to prostate cancer being more prevalent, more aggressive, and/or more frequently diagnosed in black men is unknown.

Methods: We estimated three independently developed models of prostate cancer natural history in black men and in the general population using an updated reconstruction of PSA screening, based on the National Health Interview Survey in 2005, and prostate cancer incidence from the Surveillance, Epidemiology, and End Results program in 1975–2000. Using the estimated models, we compared prostate cancer natural history in black men and in the general population.

Results: The models projected that 31–45% (range across models) of black men develop preclinical prostate cancer in their lifetime, a risk that is (relatively) 24–54% higher than in the general population. Among men who have had preclinical disease onset, black men have a similar risk of diagnosis compared with the general population, but their risk of progression to metastatic disease by the time of diagnosis is 38–75% higher than in the general population.

Conclusions: Prostate cancer incidence patterns implicate higher incidence of preclinical disease and higher risk of metastatic progression among black men. The findings suggest screening black men earlier than white men and support further research into the benefit-harm tradeoffs of more aggressive screening policies for black men.

Introduction

Prostate cancer is the most frequent cancer diagnosis and the second leading cause of cancer death in US men. Black men in the US have significantly higher prostate cancer incidence and mortality than white men, but reasons for this disparity are unclear. Some studies[1, 2] have suggested that differential access to care may partially explain the greater burden of disease in blacks, but others have concluded that differences in outcomes are more plausibly attributed to biologic differences.[3-5]

There is no question that the higher prostate cancer mortality in black men is at least partially due to their increased incidence of disease.[6] What is not known, however, is whether the higher observed incidence in black men arises from a higher risk of disease onset or faster progression to an aggressive or symptomatic state. In their multi-ethnic study of UK men, Metcalfe et al.[7] suggested that the latter is unlikely; however, they do not formally interrogate this hypothesis.

Understanding whether and how natural history might be different in black men is important because, if black men have a higher susceptibility to prostate cancer and/or a greater tendency to develop aggressive disease, it may be of value to consider different screening policies for them. This issue was raised by Powell et al.,[8] who recommended aggressive screening of black men beginning at age 40 based on a narrowing of prostate cancer survival disparities observed following the adoption of PSA screening in the US.

We previously studied the natural history of prostate cancer in the general population via statistical and computer modeling of latent disease onset and progression to clinical and metastatic states.[9-11] By calibrating the models to observed population patterns of prostate cancer incidence before and after the advent of PSA screening, we estimated the risks of critical events in disease natural history and used these results to make inferences about potential impacts of different screening policies.[12, 13]

In this article, we develop versions of our natural history models that pertain to black men and calibrate these using incidence trends in the Surveillance, Epidemiology, and End Results (SEER) program under updated PSA screening frequencies estimated specifically among the black population. We use the calibrated models to produce estimates of disease onset, progression, and diagnosis risks that pertain to the black population. We compare these risks with estimates for the general population (i.e., all races) to determine the extent to which the increased incidence among black men is explained by higher risks of disease onset, progression, or diagnosis. Finally, we use our results to motivate consideration of differential screening policies among black men.

Methods

In this section, we describe the data and three models that we use to examine evidence of differential prostate cancer natural history in black men. We also describe a test for differences in black natural history relative to the general population and quantify the models' goodness-of-fit after re-estimating key components of natural history.

PSA screening and prostate cancer incidence data

Because population-based PSA screening utilization was not tracked in real time, we retrospectively reconstructed PSA screening patterns in the US separately for black and white men in a previous study.[14] Briefly, this reconstruction used responses to the National Health Interview Survey (NHIS) in 2000 to estimate the age at first PSA test, and longitudinal claims data from the linked SEER-Medicare database to estimate the distribution of inter-screening intervals. We updated this model of PSA screening patterns using responses to the NHIS in 2005, which moderately reduced the uptake of first PSA tests in the 1990s among the oldest age groups relative to our previous analysis.[14]

We extracted prostate cancer incidence data from the SEER database before and after the introduction of PSA screening. Specifically, we extracted prostate cancer incidence for ages 50–84, years 1975–2000, SEER historic stages local-regional and distant, tumor grade well or moderately differentiated (low-grade) versus poorly differentiated or undifferentiated (high-grade), and race categories "black" or "all races." Missing information on stage, grade, and race was assumed to be missing at random and imputed as the most frequent combination of 20 logistic regression imputations using the *mice* package in R.[15]

Three models of prostate cancer natural history

We estimated three models of prostate cancer natural history using PSA screening and prostate cancer incidence data separately for black men and for all races. The three models were previously used to study effects of PSA screening on incidence and mortality trends in the general US population.[16, 17]

Briefly, the FHCRC model is a microsimulation model that links individual PSA growth and cancer progression. In this model, higher and increasing PSA levels are associated with the presence of latent cancer and shorter intervals to metastatic spread and clinical presentation. The MISCAN model is a microsimulation model

that tracks progression through combinations of cancer stages and grades. In this model, advanced stages and higher grades are associated with potentially higher screening test sensitivity and shorter intervals to clinical presentation. The UMICH model is an integrated suite of analytic models that estimates transition probabilities from earlier to later stages and from lower to higher grades during the preclinical detectable phase. In this model, a later stage at onset, a higher grade at onset, and faster progression are each associated with shorter intervals to clinical presentation. In each model, screening potentially detects latent cancer at an earlier stage and/or grade. Key differences between models are the length of the preclinical detectable phase, how much early detection improves tumor characteristics, and how both natural history and screening effects depend on age.

A framework to explain incidence disparities

Sequential estimation

We first re-estimated natural history in all races using the SEER incidence and updated PSA screening data. Then, we re-estimated natural history in black men following a systematic sequence of steps. First, we substituted PSA screening patterns for black men. Then we re-estimated components of disease natural history, each containing a specific block of parameters. The blocks of parameters governed (a) risk of disease onset and initial tumor features, (b) risks of progression to metastasis and/or high-grade disease, and (c) risk of clinical diagnosis. At each step, the re-estimation involved identifying values of the natural history parameters that allowed the models to most closely match SEER prostate cancer incidence in black men. All models proceeded in this sequential fashion until final versions of the models were obtained that re-estimated all natural history parameters for black men.

Natural history summary measures

Given the final versions of the models for black men and for all races, we summarized natural history in terms of lifetime risks of preclinical onset, clinical diagnosis, and metastatic clinical diagnosis; mean ages at these natural history events; and mean years between consecutive events.

Testing and quantifying contributions to incidence disparities

We used a likelihood ratio test to evaluate whether re-estimating components of disease natural history significantly improved the models' fits to the incidence data for black men. The likelihood uses age at diagnosis as a survival time and is fit via a customized age-period approach.[11] To calculate likelihood ratio statistics, two

likelihood functions are fit, one with and one without re-estimation of the component. While we report the likelihood ratio test results, we anticipate that, given the large sample size in the SEER registry, all tests will be highly significant at a traditional threshold. Therefore, we also report the improvement in the goodness-of-fit achieved by re-estimating components of natural history, with goodness-of-fit expressed as the sum across years of the squared difference between annual model-projected and observed age-adjusted incidence rates.

Results

Figure 1 illustrates the annual percentage of men ages 50–84 who received at least 1 PSA test by race and age group over the period 1988–2000 using responses from the NHIS in 2005. Relative to previous estimates using responses from the NHIS in 2000,[14] we find that younger men received more tests, particularly in the late 1990s, and older men received fewer tests; these differences were similar among black men and all races. The updated screening patterns indicate that, relative to the general population, modestly lower percentages of black men received at least 1 PSA test in all but the youngest ages throughout the 1990s. The greatest racial disparities in PSA testing were in the oldest ages.

Figure 1. Annual percentage of men receiving at least 1 PSA test based on the updated reconstruction of PSA screening patterns in the US.

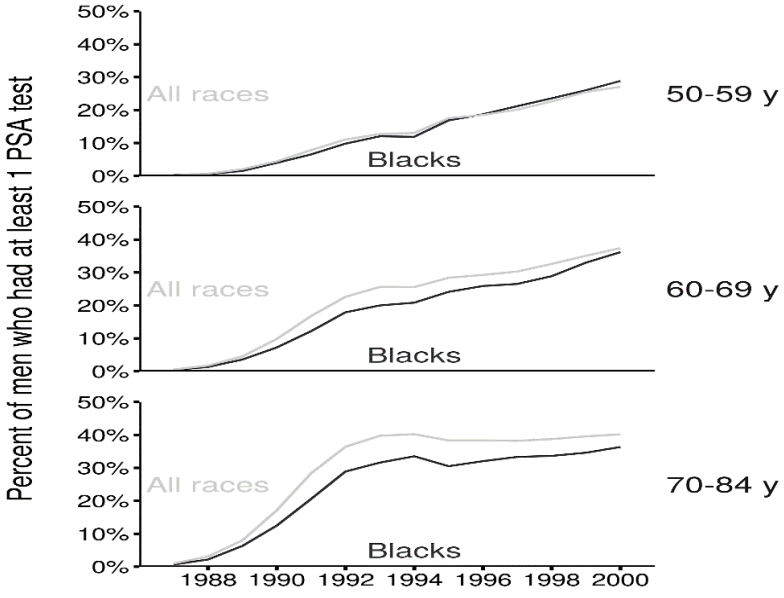
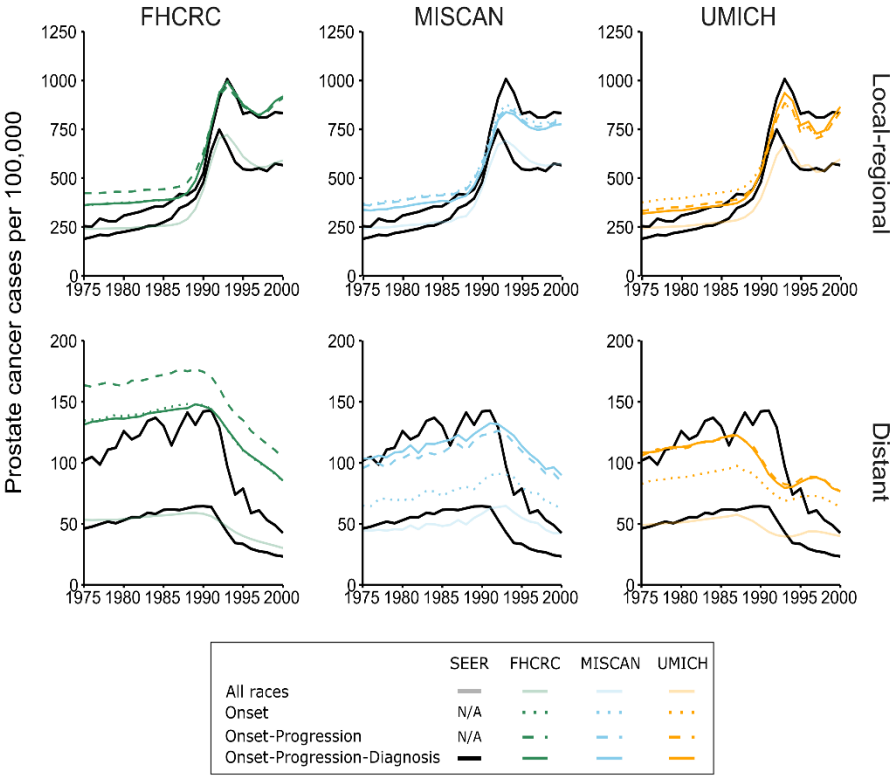


Figure 2 shows the results of re-estimating natural history for all races. The figure shows age-adjusted prostate cancer incidence rates per 100,000 men ages 50–84 reported in SEER by historic stage and corresponding model-projected incidence rates. Figure 2 also shows SEER incidence rates for black men and results of the sequential estimation of the models' natural history components. The sequential estimation found that allowing the risk of disease onset to be different for black men provided an immediate improvement in the models' fits to incidence in this population. Allowing the risk of progression to distant stage to be different produced higher distant-stage but similar local-regional stage incidence projections. And also allowing the risk of clinical diagnosis to differ in black men provided modest improvements to the fit in some cases (e.g., distant-stage incidence in the FHCRC model).

Figure 2. Age-adjusted prostate cancer incidence rates per 100,000 men ages 50–84 years for black men (black line) and all races (gray line) and corresponding projections by three models (colored lines). Model projections are based on the models estimated for all races combined with PSA screening in black men and sequentially re-estimating components of natural history to allow differential risk of onset of preclinical cancer ("Onset"), risk of progression to metastasis and/or higher grade ("Onset-Progression"), and risk of clinical diagnosis ("Onset- Progression-Diagnosis").



Improvements from re-estimating each block of natural history parameters were highly statistically significant from likelihood ratio tests (all $P < 0.0001$), and the final models' fits to stage-specific incidence were substantially improved by re-estimation of the natural history components. Table 1 shows that sums of squared differences between observed and projected age-adjusted incidence rates declined dramatically once disease onset was re-estimated, confirming the importance of disease onset risk in explaining incidence disparities. All models obtained the best fits (i.e., smallest errors) when all parameter blocks were re-estimated.

Table 1. Squared differences between age-adjusted prostate cancer incidence rates per 100,000 men ages 50–84 years for black men and corresponding model projections, summed across years 1975–2000. Model projections are based on the models estimated for all races combined with PSA screening in black men ("Black screening") and sequentially re-estimating components of natural history to allow differential risk of onset of preclinical cancer ("Onset"), risk of progression to metastasis and/or higher grade ("Onset-Progression"), and risk of clinical diagnosis ("Onset-Progression-Diagnosis").

Run	FHCRC	MISCAN	UMICH
Black screening	287160	271140	340010
Onset	41130	18340	38960
Onset-Progression	121280	23180	30360
Onset-Progression-Diagnosis	39250	18670	18010

Table 2 summarizes natural history measures among black men and for all races estimated by the three final models. In the general population, the lifetime risk of developing preclinical disease is 25–29% (range across models). In black men, however, these risks rise to 31–45%, reflecting risks that are (relatively) 24–54% higher than the general population. According to the models, the risk of clinical diagnosis in black men is 25–66% higher than the general population; the corresponding observed risk in SEER prior to the advent of PSA screening was 53% higher in black men than white men (range 42%–62% higher) over the period 1975 to 1986. Among men who have had disease onset, the risk of clinical diagnosis is comparable for blacks (39–88% across models) and all races (36–85% across models), and this translates into sojourn times from disease onset to diagnosis that are very similar for black men and the general population. However, among men with preclinical disease, the models estimate a 38–75% higher risk of metastasis before diagnosis among black men, reflecting greater risk of progression in this population.

Table 2. Natural history summary measures for black men (Blacks) and for all races (All) projected by the three models. Lifetime risks are given as percentages; mean ages and durations are given in years.

Measure	FHCRC		MISCAN		UMICH	
	Blacks	All	Blacks	All	Blacks	All
Lifetime risk of onset	45	29	31	25	37	29
Lifetime risk of clinical diagnosis	17	10	15	12	33	24
Lifetime risk of metastatic clinical diagnosis	5	2	4	3	4	2
Lifetime risk of clinical diagnosis given onset	39	36	48	47	88	85
Lifetime risk of metastatic clinical diagnosis given onset	10	6	14	10	12	7
Mean age at onset	58	60	65	68	65	66
Mean age at clinical diagnosis	73	74	72	74	80	81
Mean age at metastatic clinical diagnosis	73	75	73	76	74	74
Mean years from onset to clinical diagnosis	19	18	10	10	17	18
Mean years from onset to metastatic clinical diagnosis	16	16	12	13	16	21
Lifetime risk of PSA or clinical diagnosis*	18	11	20	17	33	24
Lifetime risk of PSA or clinical diagnosis given onset*	41	39	66	68	88	85
Mean age at PSA or clinical diagnosis*	72	73	70	72	78	80
Mean years from PSA to clinical diagnosis*	7	7	8	9	7	7

* These measures are in the presence of modeled PSA screening patterns in 1987–2000 and are included for reference.

Discussion

The observation that prostate cancer is more frequent and more lethal among black men than among white men has never been fully explained. Our study uses three models[9-11] previously calibrated to US population incidence trends. After updating the frequency of PSA testing and re-estimating these models for all races, we substituted PSA testing frequencies among black men into the models for all races and then systematically explored re-estimating individual components of disease natural history to fit disease incidence patterns in this population. The model results consistently showed that the risk of onset of a preclinical prostate cancer explains a large majority of the observed incidence disparities. Additional improvements achieved by re-estimating other natural history components produced moderate additional improvements in reproducing overall observed incidence among black men but yielded noticeable improvement in reproducing observed advanced-stage incidence. Estimation of natural history summaries based on the full race-specific

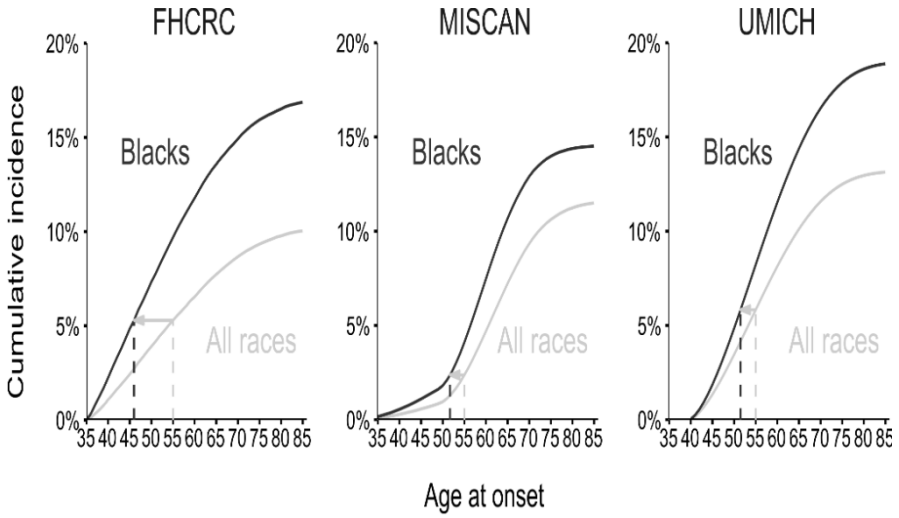
models confirmed that, in addition to the risk of onset, the risk of progression to metastatic disease before clinical diagnosis was higher among black men, but the risk of clinical diagnosis was not higher than in the general population.

Based on these results, we conclude that black men have more preclinical and progressive prostate cancer than the general population. They are more likely to develop prostate cancer at a younger age, and they are more likely to progress to a metastatic state and/or higher grade before clinical diagnosis. Their higher risk of progression agrees with a previous study based on autopsy and surgical pathology data[4] that concluded that black men had an earlier transformation to clinically significant cancer than white men. This study found similar age-specific prevalence of prostate cancer among autopsies conducted in black and white men from the Detroit metropolitan area between 1992 and 2001. The study also found evidence of more aggressive disease in radical prostatectomy specimens from black men, consistent with their markedly higher incidence of metastatic disease at diagnosis. These findings led the authors to conclude that the risk of prostate cancer initiation did not differ by race, but the risk of disease progression was higher among black men. However, similar latent prevalence and greater metastatic clinical incidence of disease among black men is in fact only possible if latent incidence is also higher in this subgroup. For, if latent incidence is similar among black men but progression is faster, this would actually lead to lower latent prevalence at autopsy. Therefore, we conclude that the prior study results are in fact consistent with our finding that the risks of latent incidence and progression are likely both higher among black men.

Our findings motivate explicitly considering more intensive screening, e.g., beginning earlier and/or screening more frequently among black men than among the general US population. Figure 3 shows the cumulative incidence of "relevant disease," i.e., disease fated to present before other-cause death. At all ages, the cumulative incidence is higher for black men than for all races. At ages 46–52 (range across models), the cumulative incidence among black men reaches the value estimated at age 55 among all races. Thus, if it is agreed that prostate cancer screening is worthwhile, and starting at age 55 is determined for the general population, our results suggest starting 3–9 years earlier for black men.

We recognize that a consensus about general population screening is still lacking. The US Preventive Services Task Force[18] recommends against routine prostate cancer screening in men of average risk, while the American Cancer Society[19] recommends shared decision making around prostate cancer screening beginning at age 50 and the American Urological Association[20] provides similar guidance with a starting age of 55. However, black men are not average risk and the benefit-harm tradeoffs of screening are likely to be different for this population.[21]

Figure 3. Cumulative incidence of onset of preclinical prostate cancer that would be clinically diagnosed in black men and all races projected by the models. Line segments show ages at which black men have incidence corresponding to levels estimated at age 55 in all races.



While starting at an earlier age is unlikely to impact overdiagnosis, other more aggressive screening policies, e.g., shortening intervals between screens or lowering the PSA threshold for biopsy referral, could increase risks of overdiagnosis.[22] A comprehensive policy development process addressing whether and how best to screen black men will have to carefully weigh benefit-harm tradeoffs of candidate policies. Understanding race-specific natural history will be a critical pre-requisite for this important work. At this point, however, our findings do support starting screening at an earlier age in black men than in the general population. Powell et al.[8] also recommend aggressive early prostate cancer PSA testing of African American men. Our work adds to theirs by providing a formal, quantitative justification for an age to initiate screening in black men.

In practice, the policy development process will require going beyond examining incidence patterns to projecting mortality in the presence and absence of screening. Since screening benefit is contingent on access to efficacious therapies, benefits of screening in different race groups may be affected by any disparities in access to treatment and any racial differences in treatment efficacy. Future work will extend the models used in this article to project the downstream outcomes of different screening policies in black men under race-specific treatment distributions and efficacies.

Limitations of this study relate primarily to modeling assumptions and data limitations. While the use of multiple models provides some sense of robustness to the specific assumptions made, all models assume that disease is progressive. Thus, none of the models explicitly includes an indolent subpopulation, although each allows heterogeneity of disease progression with some cases progressing rapidly and others slowly. The FHCRC model allows the likelihood of developing high-grade disease to vary with age but does not model grade progression; the other models allow both grade and stage progression. Other differences across models are also driven by differences in the conceptualization of onset and how the risk of onset depends on age. In the FHCRC model, onset refers to the latent incidence of disease that would be detectable by biopsy, which can occur as early as age 35, while in the MISCAN and UMICH models onset refers to the latent incidence of disease that can be detected based on elevated PSA and diagnostic workup, and this rarely occurs before age 40. Finally, the PSA screening rates used by all models are based on a retrospective reconstruction rather than a prospective tracking of prostate cancer screening dissemination in the US population.

In conclusion, this study represents the first examination of how prostate cancer natural history must differ in black men to account for racial variation in patterns of disease incidence before and after the advent of PSA screening. We use observed patterns of disease incidence and screening to learn about key events in the latent process of disease by race. Our results provide quantitative information about the prostate cancer natural history that may justify exploring different screening policies among white and black men. In pursuit of policies that will reduce disparities in disease outcomes, it may ultimately be necessary to intervene with greater intensity among population subgroups at higher risk.

References

1. Underwood W, De Monner S, Ubel P, et al. Racial/ethnic disparities in the treatment of localized/regional prostate cancer. Journal of Urology. 2004;171: 1504-1507.

2. Schwartz K, Powell IJ, Underwood W, et al. Interplay of race, socioeconomic status, and treatment on survival of patients with prostate cancer. Urology. 2009;74: 1296-1302.

3. Powell IJ. Epidemiology and pathophysiology of prostate cancer in African-American men. Journal of Urology. 2007;177: 444-449.
4. Powell IJ, Bock CH, Ruterbusch JJ, et al. Evidence supports a faster growth rate and/or earlier transformation to clinically significant prostate cancer in black than in white American men, and influences racial progression and mortality disparity. Journal of Urology. 2010;183: 1792-1796.

5. Bensen JT, Xu Z, Smith GJ, et al. Genetic polymorphism and prostate cancer aggressiveness: a case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. Prostate. 2013;73: 11-22.

6. Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. Cancer. 2012;118: 4280-4289.

7. Metcalfe C, Evans S, Ibrahim F, et al. Pathways to diagnosis for Black men and White men found to have prostate cancer: the PROCESS cohort study. British Journal of Cancer. 2008;99: 1040-1045.

8. Powell IJ, Vigneau FD, Bock CH, et al. Reducing prostate cancer racial disparity: evidence for aggressive early prostate cancer PSA testing of African American men. Cancer Epidemiology, Biomarkers and Prevention. 2014;23: 1505-1511.

9. Gulati R, Inoue L, Katcher J, et al. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics. 2010;11: 707-719.

10. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer. Journal of the National Cancer Institute. 2003;95: 868-878.

11. Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. Stat in Med. 2006;25: 2846-2866.

12. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: Model estimates of potential benefits and harms. Annals of Internal Medicine. 2013;158: 145-153.

13. Heijnsdijk EA, de Carvalho TM, Auvinen A, et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. Journal of the National Cancer Institute. 2015;107: 366.

14. Mariotto A, Etzioni R, Krapcho M, et al. Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey. Cancer. 2007;109: 1877-1886.

15. van Buuren S, Groothuis-Oudshoorn K. Multivariate imputation by chained equations in R. Journal of Statistical Software. 2011;45: 1-67.

16. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. Cancer Causes and Control. 2008;19: 175-181.

17. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118: 5955-5963.

18. Moyer VA, Force USPST. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Annals of Internal Medicine. 2012;157: 120-134.

19. Wolf AM. American Cancer Society guideline for the early detection of prostate cancer: update 2010. CA Cancer J. Clin. 2010;60: 70-98.

20. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. Journal of Urology. 2013;190: 419-426.

21. Hall L, Bullock AD, Brown AL, et al. Prostate Cancer Isn't Colorblind. New York Times, 2016.

22. Gulati R, Cheng HH, Lange PH, et al. Screening men at increased risk for prostate cancer diagnosis: Model estimates of benefits and harms. Cancer Epidemiology Biomarkers and Prevention. 2016.

Chapter Six

Estimating the risks and benefits of Active Surveillance protocols for Prostate Cancer: A microsimulation study

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk,
Harry J. de Koning.

Published: BJU Int. 2017;119(4):560-566

Abstract

Objective: To estimate the increase in prostate cancer mortality (PCM) and the reduction in overtreatment resulting from different Active Surveillance (AS) protocols, compared to treating men immediately.

Subjects and Methods: We use a microsimulation model (MISCAN-Prostate), with natural history based on ERSPC data. We estimate probabilities of referral to radical treatment while on AS, depending on disease stage, with data from Johns Hopkins AS cohort. We sample 10 million men representative of the US population and we project the effects of applying AS protocols differing by time between biopsies, compared to treating men immediately.

Results: AS with yearly follow-up biopsies for low-risk patients (≤ T2a-stage and Gleason 6) increases the probability of PCM to 2.6% (1% increase) and reduces overtreatment from 2.5% to 2.1% (18.4% reduction). With biopsies every three years after the first year, PCM increases by 2.3% and overtreatment reduces from 2.5% to 1.9% (30.3% reduction). Including intermediate-risk men (> T2a-stage or Gleason 3+4) in AS increases PCM by 2.7% and reduces overtreatment from 2.5% to 2.0% (23.1% reduction). These results may not apply to African-American men.

Conclusions: Offering AS for low-risk patients is relatively safe. Increasing the biopsy interval from yearly to up to every 3 years after the first year will significantly reduce overtreatment among low-risk men, with limited PCM risk.

Introduction

In a time of widespread debate about prostate specific antigen (PSA) screening for prostate cancer, Active Surveillance (AS) has emerged as a way to prevent the unnecessary treatment of some patients with prostate cancer, or at least, to delay the treatment of the disease [1]. The benefit of not treating a patient immediately is the avoidance of the side-effects of radical treatment.

AS consists of the carefully monitoring of men diagnosed with prostate cancer, but not yet treated, with PSA tests or repeat biopsies. What still needs to be determined is whether the benefits of avoiding side effects outweigh the risk that a patient misses his cure window by not treating immediately.

Currently, there are some AS cohort studies designed to answer this question: Prostate Cancer Research International Active Surveillance (PRIAS) [2], UCSF cohort [3], Johns Hopkins (JH) [4] or the Toronto cohort [5] among others [6,7]. However, with the exception of the Toronto cohort, their median follow-up times are shorter than 5 years. In their latest publications, very few prostate cancer deaths were reported [2-5,7].

Most AS cohorts contain only one AS protocol, usually selecting low-risk patients, with stage <T2a and Gleason Score 6 or lower (T2GS6) [6,7]. While in some cohorts follow-up biopsies occur yearly (JH [4]), in others biopsy occurs up to every 3 years after the first year (PRIAS [2]). It is not yet clear whether AS is safe for intermediate risk patients. While in the UCSF cohort [3] it was found that the 4-year treatment-free survival (TFS) did not significantly differ between low-risk and intermediate risk men, intermediate risk patients had significantly worse outcomes in the Toronto cohort [5].

Given the multitude of possible avenues for selecting and following men during AS, and the limited follow-up data, the use of modeling to evaluate the outcomes of AS protocols is necessary. Previously, Xia et al [9] compared immediate radical prostatectomy (RP) and AS for low-risk patients in a simulation study and found that AS has a modest effect on prostate cancer mortality (PCM).

In this article we use a well-validated simulation model (MISCAN) of natural history of prostate cancer, that uses JH-AS data to predict TFS, Gleason score and volume progression. We project the lifetime risk of PCM and overtreatment in the situation where AS is given to newly screen-detected low and intermediate risk men, under different follow-up biopsy intervals, and we compare these strategies with treating all men immediately.

Methods

Simulation Model

MISCAN is a microsimulation model, which simulates individual life-histories. A detailed description is available in http://cisnet.cancer.gov/prostate/profiles.html and in previously published studies [10-13].

Before detection or death, the model contains 18 health states corresponding to the combination of 3 stages (T1, T2 and T3), 3 grades (Gleason less than 7, 7, and more than 7) and whether or not cancer is metastasized. Additionally, T2-stage Gleason 6 men are classified as T2a or T2bc and Gleason 7 men are classified as 3+4 or 4+3, depending on their remaining lead-time and age group, with their respective proportions based on European Randomized Study of Screening for Prostate Cancer (ERSPC) data. Some natural history parameters were calibrated to ERSPC data (durations, transition probabilities, among others), while PSA growth parameters were calibrated to jointly to SEER incidence and ERSPC PSA distribution. [13].
In case the patient is detected outside of screening ("clinically detected") we assume that he is immediately treated. If screen-detected he can either be immediately treated or be assigned to AS.

If immediately treated, we assume an equal chance of being referred to RP or radiation therapy (RT). The prostate cancer survival without treatment is assigned at clinical detection and depends on stage and grade. It was estimated based on SEER data from the pre-PSA era (1983-1986). In order to correct for improvements on the survival not directly associated with screening or primary treatment, we add a hazard ratio for prostate cancer survival of 0.82, which was calibrated to the observed PCM in the ERSPC control (no screening) group (Model Appendix Tables 5-7).

The hazard ratios for prostate cancer survival after radical treatment equal 0.56 for RP based on [14] and 0.63 for RT (maintaining the same ratio of benefit between RP and RT from [15]). The effect of early detection is applied through an additional probability of cure which decreases exponentially with lead-time for non-metastatic cases,
Cure probability = 1 - exp(cure parameter * lead-time).

The cure parameter was calibrated to the observed PCM reduction in the ERSPC trial after 11 years of follow-up and equals -0.22. (Model Appendix Tables 5-7, Model Appendix Figure 4)

Modelling referral to treatment in AS

A patient in AS may be referred to treatment in four ways: volume progression, gleason upgrade, clinical detection and in absence of evidence of biopsy progression. If any of these events occurs then we assume that all men are treated. (Model Appendix Table 8)

Since the benefit of screening is dependent on lead-time, men who are referred to AS will experience a smaller benefit of screening, depending on how much time they are on AS. For instance, a patient with a lead-time of 10 years at screen-detection, referred to immediate treatment, will have a probability of cure as follows: $1- \exp(-0.22 * 10) = 0.89$. That is, there is an 89% probability that he is cured, and an 11% probability that he dies from prostate cancer. If the patient would choose AS and be referred to treatment 6 years later, its corresponding cure would become, $1- \exp(-0.22 * 4) = 0.59$.

TFS is defined as time from screen-detection to radical treatment. We validate TFS projected by the model, together with the number of men who experienced volume or Gleason upgrade, with data from the JH-AS study (Model Appendix Table 9). We simulated the study 100 times, by selecting patients to AS, with approximately the same age distribution and entrance criteria close to the JH cohort (maximum disease state T1 stage and GS6 and PSA $\leqslant$ 10) (Model Appendix Table 9).

Screening and Active Surveillance Policies

We sample 10 million men representative of the US age distribution based on US life tables. In the base case, we screen men yearly between ages 55 and 69, with a PSA threshold for biopsy referral (PSAt) equal to 4, biopsy compliance based on the PLCO trial and every screen-detected man is immediately treated.

We compare the outcomes of treating every man immediately, with admitting low risk patients ($\leq$ T2a, Gleason 6, PSA <10) in AS. We run a set of AS protocols where after the first year, biopsy frequency reduces to every 2, 3 or 5 years. We also project the effects of AS, with a reduced (biennial) and increased (annual, up to age 74) screening schedule. Assuming that the referral rates from AS to radical treatment, for intermediate risk men ($\leq$ T2-stage and Gleason 3+4) are similar to those of low risk men, given [7], we also project the effects of admitting low and intermediate risk men ($\leq$ T2-stage, Gleason 3+4) in AS. (Table 1)

Table 1: Active Surveillance protocols and Screening policies considered.

Basecase
1. 55-69, yearly screening, PSAt=4, No AS* (All Runs with biopsy compliance based on the observed PLCO (41%), except Run 10)
Active Surveillance
2. 55-69, yearly screening, PSAt=4, AS for ≤ T1GS6 men, with yearly biopsies.
3. 55-69, yearly screening, PSAt=4, AS for ≤ T2aGS6 men, with yearly biopsies.
4. 55-69, yearly screening, PSAt=4, AS for ≤ T2aGS6 men, with biannual biopsies after 1st year.
5. 55-69, yearly screening, PSAt=4, AS for ≤ T2aGS6 men, with triannual biopsies after 1st year.
6. 55-69, yearly screening, PSAt=4, AS for ≤ T2aGS6 men, with biopsy every 5 years after 1st year.
7. 55-69, yearly screening, PSAt=4, AS for ≤ T2aGS7 (3+4) men, with yearly biopsies. $
Sensitivity Analyses #
8. 55-69, biennial screening, PSAt=4.
9. 55-74, yearly screening, PSAt=4.
10. 55-69, yearly screening, PSAt=4 with biopsy compliance according to observed ERSPC (86%).
11. Sensitivity for Gleason Progression, probability of volume progression and leave AS increase or decrease by 20%. &
12. Null Efficacy of Treatment. §
13. Men referred only to either RP or RT. (Basecase: 50% referral for RP and the other 50% for RT) §

*PSAt denotes PSA threshold for biopsy referral. AS denotes Active Surveillance and GS denotes Gleason Score. ERSPC denotes European Randomized Study of Screening for Prostate Cancer Trial.
Each run in the sensitivity analyses is combined with run 1 (no AS), run 3 (AS with yearly biopsies) and run 5 (AS with biopsies every 3 years).
$ Also with 20% higher referral rates to treatment, as a sensitivity analysis, in Supplementary Table 3.
& See Supplementary Table 3.
§ See Supplementary Table 4.

Outcomes

The main outcome measures are the lifetime risk of PCM, treatment free life years (TFLY), which is the duration from onset of the disease until treatment and the probability of overtreatment (defined as the risk that a man is referred to radical treatment, and would not be clinically detected in absence of screening, or in other words, an overdiagnosed man who goes through radical treatment). Additionally we report the average number of years spent on AS, the probability of PCM due to entering AS and the proportion of men in AS left untreated.

Sensitivity Analyses

We run the no AS, yearly and every 3-year biopsy protocols, in combination with differential screening intensities and referral rates to treatment while in AS. We also examine the effect of no efficacy of treatment, and referring men only to either RP or RT. Since the model parameters are subject to uncertainty, we run a multivariate probabilistic sensitivity analysis including, the cure parameter, hazard ratios for treatment and the probabilities of detection in AS (Table 2). We also examined the assumption of an exponentially decreasing cure benefit, by comparing the best fit, with the fit of a linearly decreasing cure benefit. (Model Appendix Table 7).

Table 2: Overview of included uncertainty in the probabilistic sensitivity analyses

Parameter	Value	Distribution
hazard ratio improvement baseline survival	0.82	Lognormal(-0.20, 0.03) &
hazard ratio of RP*	0.56	Lognormal(-0.48,0.17) #
hazard ratio of RT*	0.63	Lognormal(-0.60, 0.19) #
cure parameter	-0.22	Beta (60.18, 213.38) &
probability of volume progression T2a	0.1	Uniform, Min=-20%, Max = +20%
probability of volume progression >T2a	0.4	Uniform, Min=-20%, Max = +20%
sensitivity Gleason progression	0.5	Uniform, Min=-20%, Max = +20%
probability leave AS without progression	0.04	Uniform, Min=-20%, Max = +20%

* Hazard ratio is relative to the survival without treatment. It is based on Bill-Axelson et al [14]. For hazard ratio of RT, we extrapolated based on the same ratio from Etzioni et al [15].
#The standard deviation was calculated such that the confidence interval in [14] would include approximately 95% of random draws.
& We first calculated how much would we need to vary the parameter until the model fit would become significantly worse, using as threshold 5% less or extra prostate cancer deaths (See Supplementary Table 2). Using this as a confidence interval for the parameter, we calculated a standard deviation such that 95% of the random draws would fall inside this confidence interval. Additionally, see Supplementary Table 2, for the comparison of the linearly decreasing fit with the exponentially decreasing fit.

Results

Active Surveillance with different entrance and follow-up protocols

Screening yearly between ages 55 and 69, with a PSAt of 4 (basecase) and treating every man immediately results in a lifetime risk of PCM of 2.6%. This strategy amounts to 7.4 treatment free life years (TFLY) per person, which contrasts with 8.7 TFLY per person (17.3% increase) if no PSA-based screening is performed and treat every men immediately. However, no screening results in a lifetime risk of PCM of 3.3% (Table 3).

On average, patients who entered AS remained untreated between 5.8 and 9.0 years depending on the screening and AS policy. If one refers patients in disease state T2aGS6 to AS then PCM increases to 2.6% (1.0% increase), TFLY from 7.4 to 7.7 and overtreatment reduces from 2.5% to 2.1% (18% reduction). About 27% of all AS men referred to AS remained untreated and the probability of dying due to AS is 1.8%.

Increasing the biopsy interval after the first year of follow-up from yearly, to two, three or five years increases PCM to about 2.6% (respectively, from 1.0% to 1.7%, 2.3% and 3.2% increase), and the proportion of men who die from prostate cancer due to AS rises from 1.8% to 3.0%, 4.1% and 5.9%, respectively. On the other hand, the probability of overtreatment reduces from 2.1% to 2.0% (25% reduction), 1.9% (30% reduction) and 1.8% (36% reduction). Average years spent on AS increases from 5.9 to 9.0.

Referring intermediate risk men to AS increases PCM to about 2.6% (2.7% increase), while overtreatment decreases from 2.5% to 2.0% (23.1% decrease). The risk of PCM due to AS is 3.6%. By contrast if we only admit low-risk men and a with biennial biopsies after the first year, the risk of PCM due to AS is only 3.0%, but with a higher overdiagnosis reduction (25.4% decrease). For univariate and multivariate sensitivity analyses see Supplementary Table 2-5 and Supplementary Figure 2.

Table 5: Active Surveillance outcomes under different screening policies and follow-up protocols.*#

	Outcomes for All Men						Outcomes for Men in AS		
	Lifetime Risk PCM (+ %)		TFLY per person (+ %)		Lifetime Risk Overtreat. (- %)		Years per person in AS	AS and not treated (%)	Risk of PCM AS$
Basecase									
55-69, yearly, PSAt=4, biopsy compliance PLCO (41%)									
No Screen	*3.26*		*8.7*	*(17.3)*	-	-	-	-	-
No AS	*2.55*		*7.4*		*2.51*		-	-	-
T1GS6, yearly biopsy	2.57	(0.6)	7.6	(2.8)	2.26	(-11.0)	6.4	26.8	1.8
T2aGS6, yearly biopsy	2.58	(1.0)	7.7	(4.2)	2.12	(-18.4)	6.0	27.1	1.8
T2aGS6, 2-year biopsy	2.60	(1.7)	7.8	(4.9)	2.00	(-25.4)	7.6	35.4	3.0
T2aGS6, 3-year biopsy	2.61	(2.3)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	4.1
T2aGS6, 5-year biopsy	2.64	(3.2)	7.9	(6.2)	1.84	(-36.4)	9.0	46.6	5.9
Sensitivity Analyses									
Increased Screening intensity (55-74, yearly, PSAt=4)									
No AS	2.38		6.7		3.88				
T2aGS6, yearly biopsy	2.42	(1.6)	7.3	(7.1)	3.09	(-25.6)	6.6	35.8	1.8
T2aGS6, 3-year biopsy	2.45	(3.1)	7.4	(9.0)	2.85	(-36.0)	8.5	46.4	3.5
Reduced Screening intensity (55-69, two-years, PSAt=4)									
No AS	2.66		7.7		2.10		-	-	-
T2aGS6, yearly biopsy	2.68	(0.7)	7.9	(2.4)	1.84	(-15.6)	6.0	32.8	2.0
T2aGS6, 3-year biopsy	2.70	(1.1)	8.0	(3.3)	1.71	(-24.3)	8.5	47.4	4.7
Biopsy Compliance ERSPC (86%)									
No AS	2.41		7.1		3.10		-	-	-
T2aGS6, yearly biopsy	2.45	(1.8)	7.6	(6.8)	2.42	(-27.0)	6.1	27.3	1.8
T2aGS6, 3-year biopsy	2.51	(2.9)	7.8	(9.4)	2.10	(-46.7)	8.6	40.7	3.0
Intermediate Risk Men									
T2GS7 (3+4), yearly	2.62	(2.7)	7.8	(5.2)	2.04	(-23.1)	5.8	23.6	3.6

* All men included in AS, had a PSA < 10 at diagnosis and a biopsy after one year. For instance, in "T2GS6, 3-year biopsy" the biopsy occurs at 1st year, 4th year and 7th year.
# PCM denotes prostate cancer mortality, TFLY denotes treatment free life years, AS stands for active surveillance, PSAt denotes PSA threshold for biopsy referral. PLCO denotes the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and ERSPC denotes European Randomized Study of Screening for Prostate Cancer Trial.
$ Risk of PCM due to AS, is the difference in PCM between the situation where every man is immediately treated and where some men are referred to AS, depending on the specific AS protocol, divided by the number of men who entered in AS.

Active Surveillance with different screening intensities

In the situation where every men is immediately treated, increasing the stopping age to 74 or increasing biopsy compliance lowers the PCM from 2.6% to 2.4% (7.4% and 6.1% decrease, respectively). On the other hand, probability of overtreatment increases from 2.5% to 3.9% (35.3% increase) or 3.1% (18.5% increase). Introducing AS, with the more intensive screening schedule seems to result in a larger effect both on overtreatment reduction and PCM increase (Table 3).

Discussion

In this study we use a novel approach to model AS, by modelling rates of volume and Gleason progression, instead of modelling durations [9] or by using a simplistic assumption of reduced treatment benefit for men in AS [17].

Introducing AS for screen-detected men results, on average, in an interval of between 5.8 and 9.0 years free of treatment, depending on the AS protocol. If we accept T2aGS6 men in AS with yearly biopsies or a biopsy every 3 years after the first year, overtreatment reduces from 2.5% to 2.1% (18.4% reduction) or 1.9% (30.3% reduction), PCM remains about 2.6% (1.0% or 2.3% increase), with a probability of dying from prostate cancer due to AS of 1.8% or 4.1%, respectively.

To put these numbers in perspective, in 2014, about 233000 men were expected to be diagnosed with prostate cancer in the US [18]. Assuming half of them are screen-detected between ages 55 and 69 (116500) and that 30% of these men are low-risk and are referred to AS (34950), our model predicts that for men with yearly biopsies or a biopsy every three years after the first year, either about 9250 or 14250 men will not be overtreated and an extra 625 or 1450 men will die of prostate cancer due to entering AS, respectively.

Our sensitivity analyses showed that these effects will become larger if the intensity of screening increases (increased stopping age or higher biopsy compliance), as more screen-detected men are classified as low-risk and have a longer lead-time. Admitting intermediate risk men in AS seems not to be as efficient as increasing the biopsy interval in AS for low-risk men.

Our modelling of AS uses a previously validated model of the natural history of prostate cancer [10-14], which is mostly based on ERSPC data [19], with US incidence validated to SEER [13,14]. By calibrating the sensitivity to Gleason Progression and the probability of detecting volume progression given that there is

an increase in stage, we are able to match the treatment free survival and the number of men experiencing volume and Gleason progression during AS to observed data in JH AS cohort.

This has some advantages relative to calibration based only on an AS cohort. First, durations and transitions between health states (which in large part determine time on AS) are based on a large randomized control trial. Second, the median follow-up of the ERSPC trial is much larger than most AS cohorts, which makes our PCM projections potentially more reliable.

On the other hand, MISCAN makes some simplifications of the AS protocol compared with the current practice in most AS cohorts. The entrance criteria for AS used in MISCAN include T-stage, GS and PSA but not the number of positive cores or PSA density. Additionally, there is some variability regarding TFS, and number of men experiencing volume and/or Gleason progression across AS cohorts [6,7].

The results in this study may not apply to African-American population [20], as the benefit of screening is estimated based on an European cohort, and the probabilities of referral to treatment while in AS are estimated based on a cohort with very few African-American men [4].

In contrast, with [9], where AS was modelled as duration from diagnosis to treatment, our approach for modelling AS allows one to project the effects of multiple AS strategies, without resorting to multiple AS cohort datasets. An additional advantage of this framework is that it allows us to jointly model screening and AS strategies. The main disadvantage of this approach is that given the difficulty of modelling directly volume progression, we need to make the assumption that volume progression can only occur, if there is an increase in T-stage in the model.

Our validation shows that MISCAN is slightly more pessimistic than the observed data in JH cohort (2 against 0 observed prostate cancer deaths). Other cohorts showed no PCM at 5 years except [21] with 1 prostate cancer death at 3.7 years of follow-up. This is likely due to the very low risk selection of patients in most AS cohorts, which contrasts with the ERSPC population used to inform natural history in MISCAN.

Importantly, we verify that a key statistic, probability of dying due to AS, which equals 1.8% in our model, is in line with previous studies where no benefit of early detection is assumed, and where AS was modelled as duration from diagnosis to treatment (Xia et al [9]: 1.2%) or as an assumption about reduced benefit (Hayes et al [17]: 2%). This rate is also comparable with the observed PCM (1.5%) in the

Toronto cohort, after 10 years of follow-up [6].

In this study we did not model quality of life. Heijnsdijk et al [22] found that introducing quality of life adjusted years (QALY's) reduces the screening benefit by 23%. Delaying treatment with AS is a way to mitigate this reduction, due to the avoidance of side-effects. For instance, Hayes et al [17] compared AS with several forms of radical treatment and found that AS gives the highest expected QALY's. Using QALY's will likely favor AS protocols that are less biopsy intensive, given the increased risk of biopsy complications [23-25].

As previously suggested [1-9], our model predicts that AS for low-risk men is relatively safe. We project the harms of benefits of several AS strategies and we find that if we increase the interval between biopsies after the first year to three years, which is close to the strategy used in the PRIAS cohort [2], overtreatment may reduce up to 30%, though with a small increase in PCM. These results apply mostly to US population of European ancestry.

References

1. Cooperberg MR, Carroll PR, and Klotz L. Active Surveillance for Prostate Cancer: Progress and Promise. J Clin Oncol 2011;29:3669-76.

2. Bul M, Zhu X, Rannikko A, et al. Radical Prostatectomy for Low-Risk Prostate CancerFollowing Initial Active Surveillance: Results From a Prospective Observational Study. Eur Urol 2012;62:195-200.

3. Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. Cancer 2008;112:2664-70.

4. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

5. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 2015;33:272-7.

6. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol 2012;62:976-83.

7. Simpkin AJ, Tilling K, Martin RM, et al. Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 2015;67:993-1005.

8. Cooperberg M, Cowan J, Hilton J, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 2011;29:228-34.

9. Xia J, Trock BJ, Cooperberg MR, et al. Prostate cancer mortality following active surveillance versus immediate radical prostatectomy. Clin Cancer Res 2012;18:5471-8.

10. Wever EM, Draisma G, Heijnsdijk EA, et al. How does early detection by screening affect disease progression?: Modelling estimated benefits in prostate cancer screening. Med Decis Making 2011;31:550-58.

11. Draisma G, Postma R, Schröder FH, et al. Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer 2006; 119:2366-71.

12. Draisma G, Boer R, Otto SJ, et al. Lead time and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 2003;95:868-78.

13. de Carvalho TM, Heijnsdijk EAM and de Koning HJ. Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136:1600-7.

14. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.

15. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012;118:5955-63.

16. Tosoian J, Mamawala M, Epstein J, et al. Intermediate and Longer-Term Outcomes From a

Prospective Active-Surveillance Program for Favorable-Risk Prostate J Clin Oncol. 2015;33:3379-85.

17. Hayes JH, Ollendorf DA, Pearson SD, et al. Active surveillance compared with initial treatment for men with low-risk prostate cancer: a decision analysis. JAMA 2010;304:2373-80.

18. American Cancer Society. Cancer Facts & Figures 2014. Atlanta, Ga: American Cancer Society; 2014.

19. 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.

20. Powell IJ. Epidemiology and pathophysiology of prostate cancer in African-American men. J Urol. 2007;177:444-49.

21. Klotz L, Zhang L, Lam A, et al. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. J Clin Oncol 2010;28:126-31.

22. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality of Life Effects of Prostate Specific Antigen Screening. N Engl J Med 2012;367:595-605.

23. Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol 2013;64:876-92.

24. Loeb S, Carter HB, Berndt SI, et al. Is repeat prostate biopsy associated with a greater risk of hospitalization? Data from SEER-Medicare. J Urol. 2013;189:867-70.

25. Nam RK, Saskin R, Lee Y, et al. Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy. J Urol 2010;183:963-8.

Supplementary Information

Supplementary Table 1: Active Surveillance outcomes under different screening policies and follow-up protocols.*# (Sensitivity Analyses)

Reduction (-) or increase (+), in percentage, relative to basecase (55-69, PSAt=4, yearly screening, No Active Surveillance)									
	Outcomes for All Men						Outcomes for Men in AS		
	Lifetime Risk PCM (+ %)		TFLY per person (+ %)		Lifetime Risk Overtreatment (- %)		Years per person in AS	AS and not treated (%)	Risk of PCM AS #
Basecase									
55-69, yearly, PSAt=4, biopsy compliance PLCO (41%)									
Stop Screen 2012	*3.09*		*8.7*	*(17.3)*	-	-	-	-	-
No AS	2.55		7.4		2.51		-	-	-
T2aGS6, yearly biopsy	2.57	(1.0)	7.7	(4.2)	2.13	(-17.9)	5.9	26.5	1.8
T2aGS6, 3-year biopsy	2.61	(2.3)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	4.1
20% lower Sensitivity									
T2aGS6, yearly biopsy	2.58	(1.3)	7.7	(4.4)	2.06	(-21.4)	6.7	30.8	2.3
T2aGS6, 3-year biopsy	2.63	(2.8)	7.9	(6.0)	1.85	(-35.3)	9.3	45.6	5.2
20% higher Sensitivity									
T2aGS6, yearly biopsy	2.58	(0.9)	7.7	(3.7)	2.16	(-16.2)	5.6	24.4	1.6
T2aGS6, 3-year biopsy	2.60	(1.9)	7.8	(5.1)	1.98	(-26.3)	7.9	36.4	3.4
T2GS7 (3+4), yearly	2.62	(2.6)	7.8	(5.2)	2.04	(-22.8)	5.8	23.4	3.5

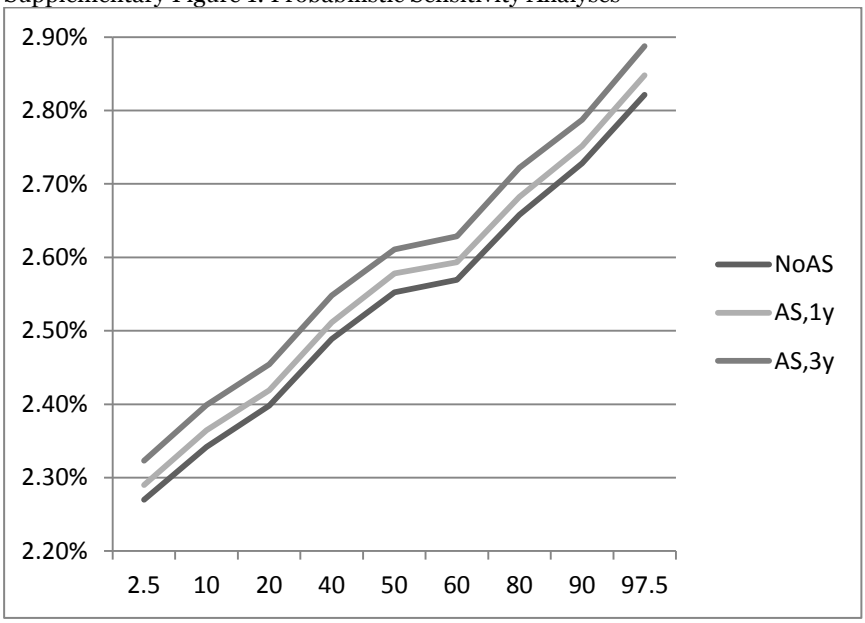
* All men included in AS, had a PSA < 10 at diagnosis and a biopsy after one year. For instance, in "T2GS6, 3-year biopsy" the biopsy occurs at 1y, 4y, 7y.
# PCM denotes prostate cancer mortality, TFLY denotes treatment free life years, AS stands for active surveillance, PSAt denotes PSA threshold for biopsy referral. PLCO denotes the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and ERSPC denotes European Randomized Study of Screening for Prostate Cancer Trial.

Supplementary Table 2: Active Surveillance outcomes under different treatment assumptions.*# (Sensitivity Analyses)

Reduction (-) or increase (+), in percentage, relative to basecase (55-69, PSAt=4, yearly screening, No Active Surveillance)									
	Outcomes for All Men						Outcomes for Men in AS		
	Lifetime Risk PCM (+ %)		TFLY per person (+ %)		Lifetime Risk Overtreatment (- %)		Years per person in AS	AS and untreated (%)	Risk of PCM due to AS #
Basecase									
55-69, yearly, PSAt=4, biopsy compliance PLCO (41%)									
Stop Screen 2012	*3.09*		*8.7*	*(17.3)*	-	-	-	-	-
No AS	*2.55*		*7.4*		*2.51*		-	-	-
T2aGS6, yearly biopsy	2.57	(1.0)	7.7	(4.2)	2.13	(-17.9)	5.9	26.5	1.8
T2aGS6, 3-year biopsy	2.61	(2.3)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	4.1
Only RP used									
No AS	2.49		*7.4*		*2.51*		-	-	-
T2aGS6, yearly biopsy	2.51	(1.0)	7.7	(4.2)	2.13	(-17.9)	5.9	26.5	1.8
T2aGS6, 3-year biopsy	2.54	(2.3)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	4.0
Only RT used									
No AS	2.62		*7.4*		*2.51*		-	-	-
T2aGS6, yearly biopsy	2.65	(1.0)	7.7	(4.2)	2.13	(-17.9)	5.9	26.5	1.9
T2aGS6, 3-year biopsy	2.68	(2.2)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	4.2
Null Efficacy of Treatment (and Screening)									
No Screening	4.06		*7.4*		*2.51*		-	-	-
No AS	4.06	(0.0)	7.7	(4.2)	2.13	(-17.9)	5.9	26.5	-
T2aGS6, 1-year biopsy	4.06	(0.0)	7.8	(5.5)	1.92	(-30.3)	8.5	40.6	0.0

* All men included in AS, had a PSA < 10 at diagnosis and a biopsy after one year. For instance, in "T2GS6, 3-year biopsy" the biopsy occurs at 1y, 4y, 7y.
# PCM denotes prostate cancer mortality, TFLY denotes treatment free life years, AS stands for active surveillance, PSAt denotes PSA threshold for biopsy referral. PLCO denotes the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial and ERSPC denotes European Randomized Study of Screening for Prostate Cancer Trial.

Supplementary Figure 1: Probabilistic Sensitivity Analyses*



*In the y-axis is the the lifetime risk of overdiagnosis, in the x-axis the percentile of the overdiagnosis distribution over a 1000 model runs. Running the model 1000 times, resulted in an approximately 95% tolerance interval of 2.27-2.82 for PCM, if every man is treated immediately. Critically it seems that the absolute effects of introducing AS on the lifetime risk of prostate cancer mortality and overtreatment do not vary much, if we change the parameter values up to about 20%.

Supplementary Table 3: Probabilistic Sensitivity Analyses: distribution of the lifetime risk of PCM and overtreatment* (%)

Lifetime Risk PCM									
Percentil	**2.5**	**10**	**20**	**40**	**Model Estim.**	**60**	**80**	**90**	**97.5**
NoAS	2.27	2.34	2.40	2.49	2.55	2.57	2.66	2.73	2.82
AS, 1y	2.29	2.36	2.42	2.51	2.58	2.59	2.68	2.75	2.85
AS, 3y	2.32	2.40	2.45	2.55	2.61	2.63	2.72	2.79	2.89
Lifetime Risk Overtreatment									
Percentil	**2.5**	**10**	**20**	**40**	**Model Estim.**	**60**	**80**	**90**	**97.5**
NoAS	2.51	2.51	2.51	2.51	2.51	2.51	2.51	2.51	2.51
AS, 1y	2.10	2.10	2.11	2.12	2.12	2.13	2.14	2.14	2.15
AS, 3y	1.88	1.89	1.90	1.92	1.92	1.93	1.95	1.95	1.96

*PCM denotes Prostate Cancer Mortality, AS denotes Active Surveillance.

Chapter Seven

Estimating the individual benefit of immediate treatment or active surveillance for prostate cancer after screen-detection in older (65+) men

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk,
Harry J. de Koning.

Published: Int J Cancer. 2016;138(10):2522-8

Abstract

A significant proportion of screen-detected men with prostate cancer is likely to be overtreated, especially in older age groups. We aim to find which groups of screen-detected older men (65+) benefit the most from Immediate Radical Treatment or Active Surveillance (AS) for prostate cancer, depending on age, screening history, health status and prostate cancer stage at detection.

We used a microsimulation model (MISCAN) of the natural history of prostate cancer based on ERSPC data. Individual life histories are simulated with US comorbidity lifetables based on a random sample of MEDICARE data. Different screening histories are simulated and we count outcomes for men screen-detected from ages 66 to 72.

For immediately treated men with low-risk disease (≤ T2a, Gleason 6) the probability of overtreatment ranges from 61% to 86% decreasing to between 37% and 46%, if they are assigned to AS. For intermediate risk men (≤ T2, Gleason 3+4) overtreatment ranges from 23% to 60%, which reduces to between 16% and 31% for AS. For high risk men (T3, or ≥ Gleason 4+3) overtreatment ranges from 11% to 51%. The disease stage at screen-detection is a critical risk factor for overtreatment.

For low risk men, AS seems to significantly reduce overtreatment at a modest cost. For intermediate risk men, the decision between immediate treatment or AS depends on age and comorbidity status. Men screen-detected in a high risk disease stage may benefit from immediate treatment even beyond age 69.

Introduction

Overdiagnosis consists of the detection of cancer that would not develop into clinical cancer in absence of screening, or eventually be life-threatening. If an overdiagnosed man is referred for radical treatment then he is considered to be overtreated. In a context of limited healthcare resources, it is crucial to identify which patients are at risk of overtreatment and therefore could be better handled by Active Surveillance (AS) or be left untreated.

Estimates of overdiagnosis in prostate cancer can vary considerably. Depending on, whether an excess incidence approach is used or, on the specific microsimulation model, estimates range between 23% and 60% of screen-detected men[1,2].

Due to these estimates, and to the controversy surrounding the magnitude of the benefit of screening in the two largest prostate cancer screening randomized control trials [3,4], the US Preventative Services Task Force recommended in 2012 against PSA screening for prostate cancer [5]. Others, like the AUA recommend shared decision making about screening between ages 55 and 69 [6].

Translating these results from the population level to an individual patient though, can be an arduous task. The probability that an individual benefits from screening and treatment can vary substantially depending on the individual person's characteristics. Namely age, health status and past screening history could have an influence on the risk of being overdiagnosed and overtreated.

Previous studies[7,8] have assessed the association between PSA, age and overdiagnosis, finding that the majority of overdiagnosed cases occurs in men older than 60. On the other hand, Wever et al [9] studied the relationship between overdiagnosis and disease stage at detection and found that for men in age group 65-69, the probability of overdiagnosis ranges from 9% to 50%. None of these studies included Active Surveillance (AS) as an option for newly screen-detected men.

AS for prostate cancer consists on the frequent monitoring of newly screen-detected men, through PSA tests and repeat biopsies. It has recently emerged as a viable alternative to immediate treatment, namely for low risk men [10]. For men older than 65, it could be an avenue to significantly reduce overtreatment. Additionally, no study has yet modelled the association between overdiagnosis and comorbidity in prostate cancer for screen-detected men. This association could have a critical impact on the estimates of [7,9], especially for men with significant comorbidities.

In this study, we use a simulation model to compute personalized estimates of overtreatment, prostate cancer mortality (PCM), number needed to treat (NNT) and life-years saved given age at screen-detection, health status, disease stage and screening history. We present results for men older than 65, which is the age group where the benefit of immediate treatment is most debatable.

Materials and Methods

Microsimulation Model

Microsimulation Screening Analysis (MISCAN) is a microsimulation model designed to study the effect of screening on incidence and prostate cancer mortality. A description is available in http://cisnet.cancer.gov/prostate/profiles.html. and in the Model Appendix.

Each of MISCAN's health states denote different disease, detection and treatment phases. We model 18 disease states, consisting of the combination of 3 stages (T1, T2, T3), 3 grades (which correspond to Gleason Score 2-6, 7 and > 7) and whether the cancer is metastasized. Additional disease states were created to model AS, including dividing Gleason 7 cancers in 3+4, 4+3 and T2GS6 (GS6 denotes Gleason Score 6 or lower) cancers between T2a and T2bc.

The transition probabilities and durations between different stages, grades until metastasized cancer and clinical diagnosis are calibrated to European Randomized Study of Screening for Prostate Cancer (ERSPC) data [11,12]. The model for PSA growth is based on [13] and is calibrated jointly to SEER incidence and ERSPC PSA distribution data [4] (Model Appendix Tables 2-4, Model Appendix Figures 2 and 3).

In an "immediate treatment" run, all screen-detected men are assigned with equal chance to either radiation therapy (RT) or radical prostatectomy (RP). In an "AS" run, all low-risk (≤ T2a Gleason 6), and intermediate-risk (Gleason 3+4, T2bc and Gleason 6) screen-detected men are assigned to AS. In the absence of treatment, a baseline survival is assigned at clinical detection, which depends on Gleason Score and is based on SEER data from the pre-PSA era. The baseline survival is further corrected by adding a hazard ratio of 0.82, which was found by calibration to the observed prostate cancer mortality in the control (no screening) group of the ERSPC [4] trial (Model Appendix Tables 5 and 6).

The hazard ratio of prostate cancer survival after treatment equals 0.56 for RP [14] or 0.63 for RT, using the same ratio as in [15]. Additionally, there is an effect of early detection: a probability of cure which decreases exponentially with lead-time for non-metastatic cases,

Cure probability = exp(cure parameter * lead-time).

The cure parameter was calibrated to the observed prostate cancer mortality reduction due to screening in the ERSPC [4] and equals -0.22. (Model Appendix Tables 5 and 6, Model Appendix Figure 4).

Active Surveillance

If a man is assigned to AS, the disease progresses as if not screen-detected. Referral to radical treatment may occur by volume progression (if there is an increase in T-stage), gleason progression, in case the patient would be clinically diagnosed or due to their own personal preference (Model Appendix Table 8). Rates of progression are dependent on the disease stage and are calibrated to Johns Hopkins cohort data [16] (Model Appendix Table 9). In the base case, we stop AS after 6 years (six biopsies) if no progression is detected, since we are modelling older men.

Comorbidity Lifetables

Three cohorts of 5 million men born in 1960 are simulated with life tables corresponding to no comorbidity, moderate comorbidity or severe comorbidity. Each comorbidity group was identified using a comorbidity index defined by conditions included in the Charlson index [17], based on a random sample of Medicare data from SEER areas. The estimation of these lifetables, corresponding life expectancies, and prevalences of each comorbidity type are extensively described in Cho et al [18]. In Vogelaar et al [19], the life expectancies corresponding to each comorbidity for several ages are shown.

For instance, moderate comorbidity includes conditions like diabetes or cerebrovascular disease, while severe comorbidity includes chronical renal failure, chronic obstructive pulmonary disease or dementia, among others. This results in an estimated prevalence at age 69 of 75% for no comorbidity, about 12% for moderate comorbidity and 13% for severe comorbidity [19].

Outcomes

We present three ages of screen-detection, 66, 69 and 72, and we consider two screening histories. In the first, the person is being screened yearly from age 55 and with a PSA threshold for biopsy referral (PSAt) of 4 until age of screen-detection. In the second, age of screen-detection is the age of first screening. These two screening histories provide a range for the outcomes that will contain most screening histories.

We follow men from screen-detection to PCM or other causes death. These are divided by age and disease stage at screen-detection, comorbidity status and screening history. The low risk group consists of men detected in stages ≤ T2aGS6. The intermediate risk group contains ≤ T2 stage, Gleason7 (3+4) men. Whereas the high risk group contains any men in either T3 disease stage and/or Gleason equal or higher than 4+3.

The main outcome measures are number needed to treat to save a life (NNT), probability of overtreatment (defined as a treated man, who is overdiagnosed) and life years (LY) saved. A man is counted as overtreated, if he is referred to radical treatment and in absence of screening, he would not be clinically diagnosed during his lifetime.

Scenario Analyses

We examine the effects of considering different AS follow-up protocols on our estimates. We vary the stop age of AS from six years after diagnosis to ten years, and we change the biopsy frequency from yearly to every three years after the first year.

Sensitivity Analyses

We run an univariate sensitivity analysis where the cure parameter and the probabilities of detection in AS decrease or increase by 20%. Additionally, we run a multivariate probabilistic sensitivity analysis, where parameters related to the benefit screening and treatment benefit are varied simultaneously (Table 1).

Table 1: Overview of included uncertainty in the probabilistic sensitivity analyses

Parameter	Value	Distribution
hazard ratio improvement baseline survival	0.82	Uniform, Min=-20%, Max = +20%
hazard ratio of RP*	0.56	Normal, Standard Deviation: 0.11*
hazard ratio of RT*	0.63	Normal, Standard Deviation: 0.12*
cure parameter	-0.22	Uniform, Min=-20%, Max = +20%

* Hazard ratio is relative to the survival without treatment. It is based on Bill-Axelson et al [14]. For hazard ratio of RT, we extrapolated based on the same ratio from Etzioni et al [15]. The standard deviation was calculated such that the confidence interval in [14] would include approximately 95% of random draws.

Results

In Table 2, we show estimates of the probability of overtreatment, PCM, and NNT, divided by age of screen-detection (66, 69, 72), comorbidity levels, prostate cancer disease stage at detection and screening history.

The prostate cancer stage at detection has a large influence both on NNT and the probability of overtreatment. For men detected in a low-risk stage, the NNT ranges from 11 to 47 and overtreatment from 61% to 86%. By contrast, men detected in a high risk disease stage have a NNT ranging from 2 to 8 and overtreatment between 12% to 51%.

The probability of overtreatment and NNT increases with age at screen-detection. For men detected at age 66 and screened once, with low risk disease and no comorbidity, the NNT increases from 11 to 15, for age 69, and to 20, for age 72. In a similar fashion, overtreatment and NNT increase with the level of comorbidity. For instance, at age 66 the probability of overtreatment for men with no comorbidity is 61% while for men with severe comorbidity is 77%.

Comorbidity and age play a special role for men screen-detected in an intermediate risk disease stage. Men with no comorbidity younger than 70 have a NNT ranging from 5 to 6 and a probability of overtreatment from 23% to 28%, which is closer to the estimates for high risk men. By contrast, men older than 70 with disease burden have a NNT between 10 and 15 and overtreatment between 41% and 52%, which is close to the predicted ranges for low risk men.

Men can also choose to delay treatment with Active Surveillance. For low risk man, this results in a large reduction of overtreatment. For instance, for a men aged 66 with no comorbidity, overtreatment decreases from 61% to 37%, with PCM increasing from 1% to 2.8%. This reduction in overtreatment becomes larger with

higher age and comorbidity, while the increase in PCM becomes smaller. For intermediate risk men, aged 66 with no comorbidity, overtreatment reduces from 23% to 16%, with PCM increasing from 5% to 10%.

In Figure 1, LY saved per screen-detected men, by disease stage, comorbidity and age are shown, for men which were yearly screened. LY saved by treatment decrease with both age and comorbidity. There are large differences in life years saved per disease stage at detection. In most cases, the amount of life years saved more than doubles if a man is detected in a higher risk disease stage than another. Referring men to AS results in a decrease in LY saved, especially for intermediate risk men.

Scenario Analyses

Increasing the biopsy interval to three years after the first year, and stopping AS after 10 years reduces overtreatment from 39.3% to 30.2% and increases PCM from 2.5% to 3.8%, compared to base case AS. By contrast, increasing the maximum time on AS from 6 to 10 years reduces PCM from 2.5% to 1.8% and increases overtreatment to 50%. (Supplemental Information Table 1).

Sensitivity Analyses

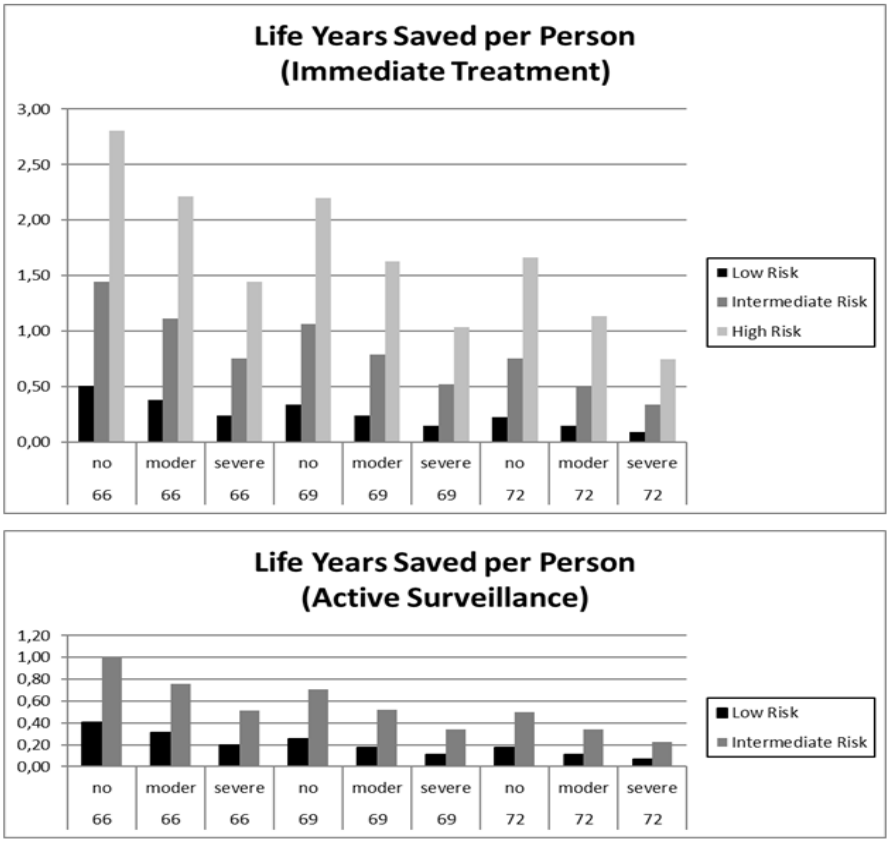
In Supplemental Information Table 2, the univariate sensitivity analyses are shown. Varying the cure rate by 20% has little effect on NNT. Changing the probabilities of detection by 20%, has a maximum effect on overtreatment of 4 percentual points. Both analyses have a modest impact on prostate cancer mortality. The probabilistic sensitivity analysis resulted in a range for NNT of men screen-detected at age 66, with low risk prostate cancer, between 10 and 14 (1^{th} and 99^{th} percentile runs). (Figure 2, Supplemental Information Figures 1-2).

Table 2: Probabilities of PCM and overtreatment at screen-detection by current age, comorbidity, screening policy and disease stage.*

		Age at detection	66			69			72		
Scr Pol	PCa Risk	Comorbid.	No	Mod	Sev	No	Mod	Sev	No	Mod	Sev
55-Age	Low	Overtreat. [IRT]	64.2	71.0	80.3	68.8	75.5	83.5	73.5	80.3	86.4
		PCM [IRT]	0.8	0.7	0.4	0.8	0.6	0.4	0.6	0.5	0.1
		Overtreat. [AS]	39.3	41.4	38.4	41.9	44.0	39.8	44.2	45.5	39.4
		PCM [AS]	2.5	1.9	1.3	2.2	1.7	1.1	1.6	1.2	0.3
		NNT #	12	15	24	16	22	35	21	31	45
	Int.	Overtreat. [IRT]	30.7	37.9	50.7	35.5	43.1	54.6	41.1	49.4	59.8
		PCM [IRT]	5.1	4.4	3.3	4.7	3.8	2.8	4.2	3.3	2.4
		Overtreat. [AS]	20.3	24.2	28.1	23.2	27.1	29.7	26.2	30.1	31.3
		PCM [AS]	10.5	8.8	6.4	9.5	7.6	5.5	8.3	6.3	4.4
		NNT	5	7	10	7	8	12	8	11	16
	High	Overtreat.	19.9	28.9	46.1	22.9	31.9	47.9	26.9	36.8	51.1
		PCM	16.5	14.9	11.4	14.9	12.8	9.6	13.5	11.3	8.4
		NNT	2	2	3	4	4	6	4	5	8
1x	Low	Overtreat. [IRT]	60.6	67.6	77.1	65.7	72.8	81.4	71.0	78.3	84.8
		PCM [IRT]	1.0	0.8	0.5	0.9	0.7	0.5	0.9	0.7	0.5
		Overtreat. [AS]	36.7	39.4	36.8	39.8	42.4	38.1	41.8	43.9	38.2
		PCM [AS]	2.8	2.4	1.6	2.3	1.67	1.2	2.0	1.5	1.0
		NNT	11	14	22	15	20	32	20	29	47
	Int.	Overtreat. [IRT]	22.7	29.5	41.7	28.2	35.5	47.4	33.2	40.8	51.5
		PCM [IRT]	5.1	4.6	3.3	5.2	4.4	3.0	4.6	3.7	2.8
		Overtreat. [AS]	15.9	20.1	24.2	19.2	23.5	26.6	22.6	26.6	28.4
		PCM [AS]	10.5	9.0	6.8	9.1	7.6	5.4	8.4	6.6	4.8
		NNT	5	6	8	6	8	11	7	10	15
	High	Overtreat.	11.5	17.3	32.8	13.9	21.4	35.8	18.6	27.0	41.4
		PCM	18.1	16.3	12.6	16.3	13.8	10.7	16.2	13.7	10.4
		NNT	3	4	5	3	4	6	4	5	7

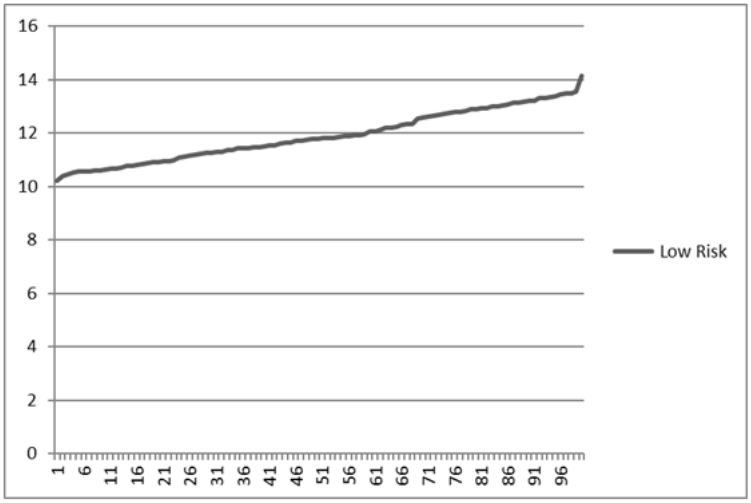
*Screen-Policies include annual screening starting at age 55 and stop at the age of screen-detection (55-Age), and one time screening at the age of screen-detection (Once). Disease Risk at screen-detection is divided in Low (≤ T2aGS6), Intermediate (≥ T2aGS6, Gleason 3+4) and High (either Gleason 4+3 or higher, or stage T3 or higher). The outcome measures are probability of overtreatment (Overtreat), probability of prostate cancer mortality (PCM) and number needed to treat to save one life (NNT). NNT is measured as the inverse of the absolute risk difference of PCM between treated and untreated men. Treatment modalities are Immediate Radical Treatment (IRT), which comprises Radical Prostatectomy (RP) and Radiation Therapy (RT) or Active Surveillance (AS). Comorbidity categories include no comorbidity ("No"), moderate comorbidity ("Mod") and severe comorbidity ("Sev").

Figure 1: Life Years Saved by treated person presented by age, comorbidity status and disease stage at detection.*#



* Disease Risk at screen-detection is divided in Low Risk (≤ T2aGS6), Intermediate Risk (> T2aGS6, Gleason 3+4) and High Risk (either Gleason ≥ 4+3, or stage ≥ T3).
#Life years saved (LY) is the difference between the life years per screen-detected person with radical treatment or active surveillance and no treatment.

Figure 2: Distribution of the NNT for low risk men, aged 66 and no comorbidity. *



*In the x-axis the percentile of the NNT distribution is shown, and in the y-axis the corresponding NNT. See Table 1 for details on included uncertainty and Supplemental Information Figures 1 and 2, for intermediate and high risk men.

Discussion

Individual characteristics, like age, comorbidity status and disease stage at detection have a decisive effect in the risk that a man does not benefit from prostate cancer screening, AS or treatment. The main strength of this study is that the effects of these individual factors are, as far as we know, for the first time quantified together when deciding between immediate treatment, AS and no treatment.

If we consider a man screened yearly since age 55 and screen-detected at age 66 with no comorbidity and in a low risk disease stage then the probability of overtreatment is 64% and NNT is 12. If this person was in a high risk disease stage this probability would decrease to 20% and NNT to 5. By contrast, if he had a severe comorbidity, NNT increases to 24 and the probability of overtreatment to 80%. If he was detected at age 69 his NNT would be 16 and probability of overtreatment would equal 69%.

Men screen-detected in a high risk disease stage have a relatively low NNT, probability of overtreatment and the largest amount of LY saved by screening. Even beyond the 55-69 age group there seems to be a significant benefit on treating these men. The same holds for intermediate risk men, younger than 70 and no comorbidities.

By contrast, for men detected in a low risk disease stage the probability of overtreatment and NNT are large. Offering AS to low-risk men older than 65 seems to be an attractive alternative to immediate treatment. Our projections show that overtreatment can be sharply reduced (> 30%), at the cost of a small PCM increase. For men older than 70 or with significant comorbidity these gains are even more pronounced.

Currently, most ongoing observational cohorts of AS select these men and so far the results seem promising with few or none prostate cancer related deaths [20-22] However, with the exception of Klotz et al [23], the median follow-up is small (< 5 years). There is some evidence in favor of including intermediate risk men in AS [24]. We find that for intermediate risk men, the overtreatment reduction and PCM increase due to AS are proportionally similar to low risk men. However, this means for intermediate risk men, aged 66 and with no comorbidity that PCM increases from 5% to 11%, which many could consider to be a bar too high. By contrast, for men aged 72 with severe comorbidity, PCM increases only from 2.8% to 4.8%, which means that personal characteristics are important when choosing between immediate treatment and AS for intermediate risk men.

Running a probabilistic sensitivity analysis with all parameters in the model would be too computationally expensive. Therefore we run an univariate sensitivity analysis for the cure parameter and the probabilities of detection in AS and a multivariate probabilistic sensitivity analysis with parameters related to the benefit of treatment and early detection only. Through this analysis we show that our results are relatively robust, since varying the values of the parameters by 20% would have a small impact in the projections.

In this study we had to limit our analyses to men older than 65. This is the case since the comorbidity lifetables were based on Medicare data. While this can be considered a serious limitation, it was already shown that a large proportion of overdiagnosis occurs in men older than 60 [7-9].

Compared to Gulati et al [7], our study tries to address some of the criticisms given in Freidlin and Korn [25]. For instance, they refer that the new USPSTF recommendation, would have a profound impact on screening patterns, and hence, on their estimates. We show that considering once in a lifetime screening (the most limited possible screening policy), changes the estimates of overtreatment and NNT, however the size of the change would have a negligible impact on any recommendation.

This model is a simplification of the natural history of prostate cancer. The comorbidity lifetables do not consider the possibility of interactions between comorbidity and the effect of treatment or cancer biology. Additionally, there is considerable uncertainty regarding the benefit of screening. Nonetheless, our results seem to be consistent with previous literature [7-9].

In this study we did not model quality of life. It was previously shown [26], that the side effects of treatment could decrease the benefit of PSA testing up to 23%. Extrapolating this result to this study would likely mean that the QALY's saved will be of a smaller magnitude than the LY saved shown in Figure 1.

We did not take into account the PSA level at screen-detection. Gulati et al [7] found that men at lower PSA levels at detection are at a higher risk of being overtreated. Extrapolating this result to this study would likely translate to lower NNT for men with high PSA values at detection.

Men screen-detected in a low risk disease stage (≤ T2aGS6) and their physicians should carefully consider the potential harms before deciding on aggressive treatment, as they are at great risk of overtreatment. A large majority of them could be safely handled with AS, and then, about a third will avoid treatment. By contrast, those in a high risk disease stage (T3 or ≥ Gleason 4+3) have a relatively low chance of overtreatment, and unless they have significant morbidities they are likely to benefit from immediate radical treatment even beyond age 69.

Our findings highlight the importance of taking several risk factors into account, when making treatment decisions. We find that the pivotal variable is disease stage at screen-detection. However, age and comorbidity should also play a significant role on the decision to do radical treatment, active surveillance or watchful waiting, especially for intermediate risk men.

References

1. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. J Natl Cancer Inst 2009; 101:374-83.

2. Welch HG, Black WC. Overdiagnosis in Cancer. J Natl Cancer Inst 2010;102:605–13.

3. Andriole GL, Crawford ED, Grubb RL, et al. Mortality Results after 13 years of Follow-up. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med 2009; 360:1310-19.

4. 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.

5. Moyer VA; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2012;157:120-34.

6. Ballentine Carter H, Albertsen PC, Barry MJ, et al. Early Detection of prostate cancer: AUA guideline. American Urological Association 2013. http://www.auanet.org/education/guidelines/prostate-cancer-detection.cfm , Accessed February 2014.

7. Gulati R, Inoue LY, Gore JL, et al. R. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst 2014;106:djt367.

8. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC Med 2014;12:26. doi: 10.1186/1741-7015-12-26.

9. Wever EM, Hugosson J, Heijnsdijk EA, et al. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. Br J Cancer 2012;107:778-84.

10. Cooperberg MR, Carroll PR, and Klotz L. Active Surveillance for Prostate Cancer: Progress and Promise. J Clin Oncol 2011;29:3669-76.

11. Draisma G, Postma R, Schröder FH, et al. Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer 2006;119:2366-71.

12. Draisma G, Boer R, Otto SJ, et al. Lead time and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 2003;95:868-78.

13. de Carvalho TM, Heijnsdijk EAM, de Koning HJ. Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136:1600-7.

14. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.

15. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012;118:5955-63.

16. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

17. Charlson M, Szatrowski TP, Peterson J, et al. Validation of a combined comorbidity index. J Clin Epidemiol 1994;47:1245-51.

18. Cho H, Klabunde CN, Yabroff KR, et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. Ann Intern Med 2013;159:667-76.

19. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, et al. Personalizing Age of Cancer Screening Cessation Based on Comorbidity: Model estimates of harms and benefits. Ann Intern Med 2014;161:104-12.

20. Bangma CH, Valdagni R, Carroll PR, et al. Active surveillance for low-risk prostate cancer: developments to date. Eur Urol 2015;67:646-8.

21. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol 2012;62:976-83.

22. Simpkin AJ, Tilling K, Martin RM, et al. Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 2015;67:993-1005.

23. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 2015;33:272-7.

24. Cooperberg M, Cowan J, Hilton J,et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 2011;29:228-34.

25. Freidlin B, Korn EL. A model too far. J Natl Cancer Inst 2014;106:djt368.

26. Heijnsdijk EA, Wever EM, Auvinen A, et al. Quality of Life Effects of Prostate Specific Antigen Screening. N Engl J Med 2012;367:595-605.

27. Bazaraa MS, Sherali HD, Shetty CM. Nonlinear programming: theory and algorithms. New York (NY): John Wiley and Sons; 1993. p. 353.

28. Wever EM, Draisma G, Heijnsdijk EA, et al. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst. 2010 Mar 3;102(5):352-5.

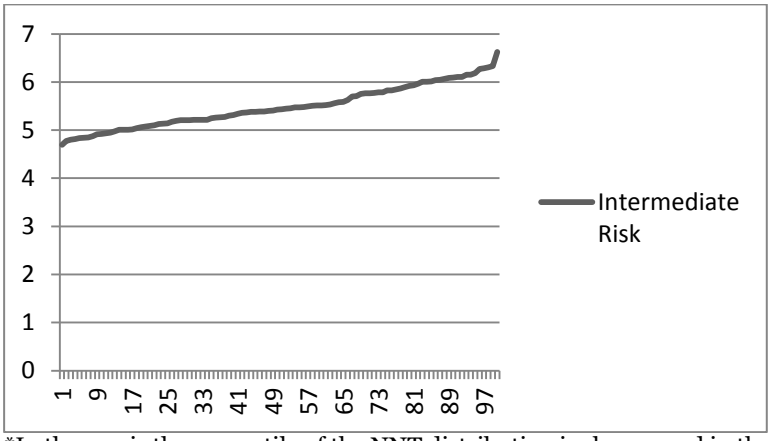
29. Inoue LY, Etzioni R, Morrell C, et al. Modeling Disease Progression with Longitudinal Markers. J Am Stat Assoc. 2008;103:259-70.

30. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.

Supplemental Information

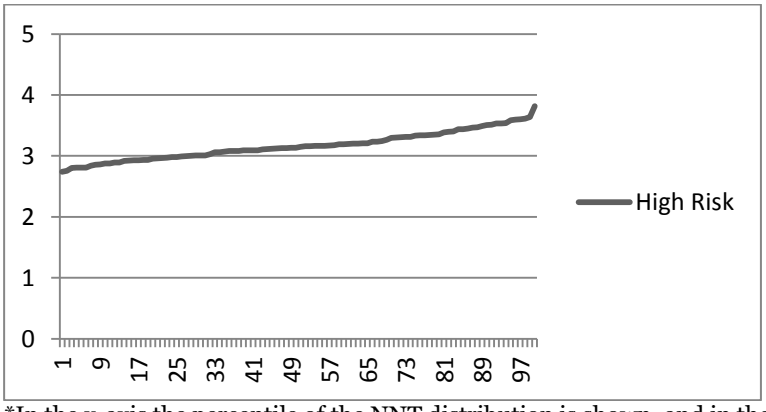
Sensitivity Analyses

Figure 1: Distribution of the NNT for intermediate risk men. *



*In the x-axis the percentile of the NNT distribution is shown, and in the y-axis the corresponding NNT. See Table 2 for details on included uncertainty.

Figure 2: Distribution of the NNT for high risk men, aged 66 and no comorbidity.*



*In the x-axis the percentile of the NNT distribution is shown, and in the y-axis the corresponding NNT. See Table 2 for details on included uncertainty.

Table 1: Probabilities of PCM and overtreatment at screen-detection for men aged 66 screened once, for different AS protocols.*

AS protocol		AS (basecase)			10 years			10 years, with triennial biopsies		
Disease Risk	Comorbid.	No	Mod	Sev	No	Mod	Sev	No	Mod	Sev
Low	Overtreat.	39.30	41.40	38.41	50.04	51.11	45.32	30.21	30.25	25.77
	PCM	2.54	1.93	1.33	1.82	1.39	0.94	3.83	2.91	1.97
Interm	Overtreat.	20.34	24.24	28.13	24.32	27.86	30.78	7.74	11.31	18.65
	PCM	10.49	8.82	6.42	9.61	8.15	5.97	12.31	10.32	7.41

*Screen-Policies include annual screening starting at age 55 up to the age of screen-detection of 66. Disease Risk at screen-detection is divided in Low (≤ T2aGS6), Intermediate (≥ T2aGS6, Gleason 3+4) Risk. Active Surveillance (AS) is only offered to low and intermediate risk men in all protocols. “AS (basecase)” denotes 6 years in in Active Surveillance with annual biopsies. “10 years” denotes 10 years in AS with annual biopsies. “10 years, with triennial biopsies” denotes a biopsy follow-up protocol where after the first year, there is a biopsy every three years. Comorbidity categories include no comorbidity (“No”), moderate comorbidity (“Mod”) and severe comorbidity (“Sev”). PCM denotes prostate cancer mortality.

Table 2: Univariate Sensitivity Analyses*

Age at detection		Basecase			20% Less Sensitivity			20% More Sensitivity		
PCa Risk	Comorbidity	No	Mod	Sev	No	Mod	Sev	No	Mod	Sev
Low	Overtreat. [IRT]	64.2	71.0	80.3						
	PCM [IRT]	0.8	0.7	0.4						
	Overtreat. [AS]	39.3	41.4	38.4	35.2	37.0	34.1	42.5	44.8	42.1
	PCM [AS]	2.5	1.9	1.3	3.0	2.3	1.6	2.3	1.7	1.2
	NNT	12	15	24						
Interm.	Overtreat. [IRT]	30.7	37.9	50.7						
	PCM [IRT]	5.1	4.4	3.3						
	Overtreat. [AS]	20.3	24.2	28.1	19.0	22.7	26.4	21.3	25.4	29.4
	PCM [AS]	10.5	8.8	6.4	11.1	9.3	6.7	10.1	8.5	6.2
	NNT	5	7	10						
					20% Less Cure			20% More Cure		
	Comorbid.	No	Mod	Sev	No	Mod	Sev	No	Mod	Sev
Low	Overtreat. [IRT]	64.2	71.0	80.3						
	PCM [IRT]	0.8	0.7	0.4	1.2	1.0	0.6	0.6	0.5	0.3
	Overtreat. [AS]	39.3	41.4	38.4						
	PCM [AS]	2.5	1.9	1.3	2.9	2.2	1.5	2.4	1.8	1.2
	NNT	12	15	24	12	16	25	12	15	23
Interm.	Overtreat. [IRT]	30.7	37.9	50.7						
	PCM [IRT]	5.1	4.4	3.3	6.4	5.4	4.0	4.3	3.7	2.8
	Overtreat. [AS]	20.3	24.2	28.1						
	PCM [AS]	10.5	8.8	6.4	11.3	9.4	6.8	9.9	8.3	6.1
	NNT	5	7	10	6	7	10	5	6	9
High	Overtreat.	19.9	28.9	46.1						
	PCM	16.5	14.9	11.4	18.9	17.0	12.8	14.5	13.2	10.1
	NNT	2	2	3	3	4	6	3	4	5

* First half of the table concerns sensitivity analyses about probabilities of detection in Active Surveillance (AS), therefore only AS estimates are affected. Lower half of the table concerns the cure rate and therefore only prostate cancer mortality (PCM) and number needed to treat to save a life (NNT) are affected. Screen-Policies include annual screening starting at age 55 up to the age of screen-detection of 66. Disease Risk at screen-detection is divided in Low (≤ T2aGS6), Intermediate (≥ T2aGS6, Gleason 3+4) and High (≥ Gleason 4+3 or ≥ T3). Active Surveillance (AS) is only offered to low and intermediate risk men in all protocols. Comorbidity categories include no comorbidity (“No”), moderate comorbidity (“Mod”) and severe comorbidity (“Sev”).

Chapter Eight

Cost-Effectiveness of different Active Surveillance protocols compared to Immediate Treatment for prostate cancer: A modelling study.

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk,
Harry J. de Koning.

Submitted

Abstract

Purpose: Active Surveillance (AS) for prostate cancer is a way to decrease overtreatment due to prostate specific antigen (PSA) based screening, among low-risk men. We estimate the cost-effectiveness of different Active Surveillance protocols compared to immediate treatment.

Methods: We use a microsimulation model of screening and AS, based on data from ERSPC and SEER, for the natural history of prostate cancer and Johns Hopkins AS cohort data to inform probabilities of referral to treatment during AS. We simulate a cohort of 10 million men based on US lifetables. We project the lifetime costs and effects of multiple screening and AS protocols and determine their cost-effectiveness.

Results: Quadrennial screening between ages 55-69 with AS for low-risk men and yearly biopsies or triannual biopsies resulted in an incremental cost per QALY of $51,918 or $69,380, respectively. Most policies where screening was followed by immediate treatment were dominated. In most sensitivity analyses, we found a policy where cost per QALY remained below $100,000.

Conclusions: AS is more cost-effective than immediate treatment. Within the AS protocols considered, admitting only low-risk men with triannual biopsies seems to be more efficient, however the differences between AS protocols are small. Limited screening combined with AS could be cost-effective at a $100,000 threshold.

Introduction

Frequent PSA-based screening in the US has led to concerns that a substantial proportion of screen-detected men may be overdiagnosed and overtreated. Active Surveillance (AS) for prostate cancer consists in the monitoring of newly diagnosed and not yet treated men through PSA tests and/or repeat biopsies.

The main goal of AS is to delay or avoid the treatment of patients who will likely derive no benefit from immediate treatment, and hence, to significantly reduce the harms of prostate cancer screening, while preserving the prostate cancer mortality (PCM) reduction.

There are currently several cohorts following the outcomes of men on AS: Prostate Cancer Research International Active Surveillance (PRIAS) [1], University of California at San Francisco [UCSF] cohort [2], Johns Hopkins [JH] [3], the Toronto cohort [4] among others [5,6]. However, with the exception of the Toronto cohort, their median follow-up times are shorter or equal to 5 years, which is not long enough to establish the long term effects [1-6].

Currently, the rates of prostate cancer specific mortality in AS are relatively low and there is an emerging consensus that AS is safe for low-risk (T-stage T2a, Gleason 6) patients [7,8]. Some cohorts include intermediate risk patients (Gleason 3+4) though with mixed outcomes [2,4].

Given the limited nature of the follow-up data of most AS cohorts, computer modelling was used to make projections of the potential effects of AS on PCM and overtreatment. Previous modelling studies [9,10] found that compared to immediate treatment, AS results in a modest increase in PCM.

However, a critical aspect missing from these studies is not to consider quality of life and costs associated with AS, when comparing with immediate treatment. For instance, Heijnsdijk et al [11] found that considering quality of life will decrease the benefits of screening by 23%, which suggests that including quality-adjusted life years (QALYs) will have a major impact on the projected harms and benefits of immediate treatment when compared to different AS protocols. Previous studies [12, 13] found that watchful waiting and AS are more cost-effective than immediate treatment, but without explicitly modelling the AS protocol.

In this study we present estimates of the cost-effectiveness of several screening policies combined with immediate treatment against screening combined with AS for low-risk men. Additionally we study whether AS protocols could be improved by admitting intermediate risk men, or by having a lower biopsy frequency.

Methods

Simulation Model

MISCAN is a microsimulation model designed to evaluate the effects of prostate cancer screening. A detailed description is available in [10,14], http://cisnet.cancer.gov/prostate/profiles.html and in previously published studies [15-17].

The natural history model contains 18 health states corresponding to the combination of 3 stages (T1, T2 and T3), 3 grades (Gleason < 7, 7, > 7) and whether or not cancer is metastasized. Additional states were created to model AS. Men in T2-stage, Gleason 6 are classified as T2a or T2bc and Gleason 7 men as 3+4 or 4+3, depending on their remaining lead-time and age group. Initially, natural history parameters, which include onset of the disease, durations and transition probabilities between health states were calibrated to observed ERSPC incidence data [15,16]. This model was adapted to the US situation by adding an extra hazard of clinical diagnosis and obtaining US-specific estimates for other parameters, using SEER data.

The prostate cancer survival without treatment is assigned at clinical detection and depends on Gleason score. It was estimated based on SEER data from the pre-PSA era (1983-1986). In order to correct for improvements on survival not directly associated with screening or primary treatment, we add a hazard ratio for prostate cancer survival of 0.82, which was calibrated to the observed PCM in the ERSPC control (no screening) group [14].

In case the patient is screen-detected and referred to AS, natural history progresses as if he was not screen-detected. A patient may exit AS due to Gleason or volume progression (which is assumed can only occur after an increase in stage) at each biopsy round, due to personal preference or if he would be clinically detected at the time. The probabilities of referral to radical treatment are estimated based on JH-AS observed treatment-free survival data, with the rate of disease progression based on our natural history model. For intermediate risk men we assume that the probabilities of referral to treatment, given progression are similar to low-risk men. (Model Appendix Tables 8 and 9).

The hazard ratios for prostate cancer survival after radical treatment equal 0.56 for radical prostatectomy (RP) based on Bill-Axelsson et al [18] and 0.63 for radiation therapy (RT) (maintaining the same ratio of benefit between RP and RT from [19]). The effect of screening is dependent on the remaining lead-time for non-metastatic cases,
Cure probability = exp(cure parameter * lead-time).

The cure parameter was calibrated to the observed PCM reduction in the ERSPC trial after 11 years of follow-up and equals -0.22 [14] (Model Appendix Table 5 and 6).

Screening and AS protocols

We simulate a cohort of 10 million men, aged 55 in 2015, based on US lifetables. In the basecase, men are screened between 55-69 (with an attendance rate of 90%), with PSA threshold for biopsy referral of 4, biopsy compliance equal to 41%, based on the PLCO trial [20], and every man is immediately treated.

We compare the immediate treatment situation with assigning low risk men (T2a-Stage, Gleason 6, PSA<10) to AS, followed with yearly biopsies. Other AS protocols, also include intermediate risk men or a triannual interval between biopsies, after the first year. We assume all men classified as low-risk are selected for AS. The biopsy compliance during AS is based on PRIAS observed biopsy compliance (See ref. 21 and Model Appendix Table 10).

Additionally, we also examine the effect of changing the intensity of screening, by combining AS with screening strategies with an earlier stopping age (61) or lower screening frequency (2 and 4-year).

Quality of Life and Costs

QALYs were calculated by using utility estimates ranging between 0 (death or worst imaginable health) and 1 (full health). Utility estimates and durations concerning all stages of early detection are based on [11, 22, 23]. In particular, the utility for the post-recovery from radical treatment was calculated based on [22,23] (Model Appendix Table 11). The costs of screening are based on [11]. The cost of immediate treatment is an estimate from another simulation model [24]. Costs of palliative therapy are based on [25]. Costs do not include indirect costs (except in sensitivity analysis). (Table 1)

Table 1: Utilities, durations and costs of screening and treatment*

Event	Utility	Duration (Years)	Cost (US dollars)
Screening	0.99	0.02	151
Biopsy	0.90	0.06	743
Cancer Diagnosis	0.80	0.08	-
RT , < 2 months	0.73	0.16	23,565#
RT , 2-12 months	0.78	0.84	
RP , < 2months	0.67	0.16	16,946#
RP, 2-12 months	0.77	0.84	
AS (Surveillance Costs)§	-	6	245 (per year)
Post-Recovery	0.95**	9	
Palliative Therapy	0.60	2.5	48,472 &
Terminal Illness	0.40	0.5	

* All utilities and durations based on [11].
# Cost of radical prostatectomy (RP) and radiation therapy (RT) includes surveillance costs [24].
§ Surveillance costs of AS, include 4 PSA tests and one visit to the doctor per year, for 6 years. These costs do not include yet the cost of biopsy, as this depends on the AS protocol.
& Based on [25].
** Based on data from [11, 22, 23]. For calculation, see Supplementary Table 3.

Outcomes

We estimate the cost-effectiveness of each AS protocol and immediate treatment. The main outcome is the incremental cost effectiveness ratio (ICER). The average cost per QALY gained is relative to the situation with no screening and assuming every clinically detected man is treated immediately. We also show cancers diagnosed and overdiagnosed, prostate cancer mortality and life years (LYs), and the overall cost of the screening program.

Sensitivity Analyses

In order to assess the effect of uncertainty around the parameter estimates on the outcomes, several multivariate sensitivity analysis are performed including the utility and cost estimates for each event, the parameters of the model related to treatment benefit and referral to treatment while in AS, the benefit due to early detection, and the effect of discounting (Table 2).

Table 2: Overview of included uncertainty in the multivariate sensitivity analyses*

Parameter	Value	Range
Cure Parameter		
cure parameter	-0.22	Min=-20%, Max = +20%
Hazard Ratio Treatment and baseline Survival		
hazard ratio improvement baseline survival	0.82	Min=-20%, Max = +20%
hazard ratio of RP	0.56	0.41, 0.77 #
hazard ratio of RT	0.63	0.46, 0.87 #
Active Surveillance Parameters		
sensitivity to Gleason Progression	0.4	Min=-20%, Max = +20%
probability detection volume upgrade (T2a)	0.1	Min=-20%, Max = +20%
probability detection volume upgrade (>T2a)	0.5	Min=-20%, Max = +20%
probability referral to treatment without progression	0.04	Min=-20%, Max = +20%
Utilities (Favourable , Unfavourable) &		
Screening	0.99	1.00, 0.98
Biopsy	0.90	0.87, 0.94
cancer diagnosis	0.80	0.85, 0.75
RT , < 2 months	0.73	0.75, 0.71
RT , 2-12 months	0.78	0.88, 0.68
RP , < 2months	0.67	0.78, 0.56
RP, 2-12 months	0.77	0.84, 0.70
post-recovery	0.95§	0.93, 0.97
palliative therapy	0.60	0.24, 0.86
terminal illness	0.40	0.24, 0.56

* In two additional separate analyses, all costs are varied by more and less 50% and cost-effectiveness is recalculated with 0% and 6% discount. RP denotes radical prostatectomy, RT denotes radiation therapy.
# Confidence interval for hazard ratio of RP, is the observed in [18]. The confidence interval of RT was extrapolated using the same ratio as in [19].
& Adapted from [11]. "Favorable/Unfavorable" refers to whether this utility gives a higher/lower QALY gained by screening, respectively.
§ Based on [11, 22 and 23]. See Supplement Table 3 for details.

Results

Effects

In Table 3, the effects of screening yearly, followed by immediate radical treatment or AS are shown. 158 cancers per 1000 screened men were diagnosed of which 53 were screen-detected, resulting in 23 overdiagnosed cancers and in a prostate cancer mortality reduction of 23%. We estimated, at a 3% discount rate, that 30 life years were saved by screening, but adjusting for quality of life, this number reduced to 17 (See Supplementary Information Table 1 for the undiscounted values).

By referring all low-risk patients to AS, life years gained reduced to 28, however QALYs increased to 18. Selecting intermediate risk men for AS resulted on 16 QALY's gained. Stopping screening at age 61 resulted on 12 QALY's gained, both with immediate treatment and AS. Reducing the frequency of screening to every 2 years resulted in about 14 to 15 QALY's gained depending on the AS protocol, or 10 to 11 QALY's gained, for quadrennial screening.

Costs

The cost of screening and treatment, yearly, between ages 55 and 69, relative to the no screening situation was about $1.8 million for 1000 screened men, for immediate treatment and $1.7 million AS with yearly biopsies. If we stop screening at age 61 or reduce the frequency to every 2 or 4 years the cost becomes, respectively, $0.9 million, $1.0 million or $0.6 million per 1000 screened men.

Average Cost per QALY

The cost per QALY gained of screening men yearly between 55 and 69 and treating every man immediately was about $103,037. By referring low-risk men to AS, with yearly or 3-year biopsies this reduced, respectively, to $91,979 or $91,654. Reducing the number of screens by stopping screening at age 61, or screen every 2 or 4 years resulted in a cost per QALY of $75,510 or $73,590 or $55,673, respectively.

Incremental Cost-Effectiveness

Of all the screening and AS policies considered, we determined which were the most efficient based on their incremental cost-effectiveness ratio. Most policies where screening was followed by immediate treatment or where screening was followed with AS for low and intermediate risk men were dominated, that is, they were more expensive and resulted on less QALYs gained.
Screening between ages 55 and 69 every four years and offering AS for low-risk men, with yearly or triannual biopsies resulted in an incremental cost per QALY lower than the $100,000 threshold.

Table 4: Costs and Effects of immediate radical treatment (IRT) compared to Active Surveillance (AS), per screening intensity.*

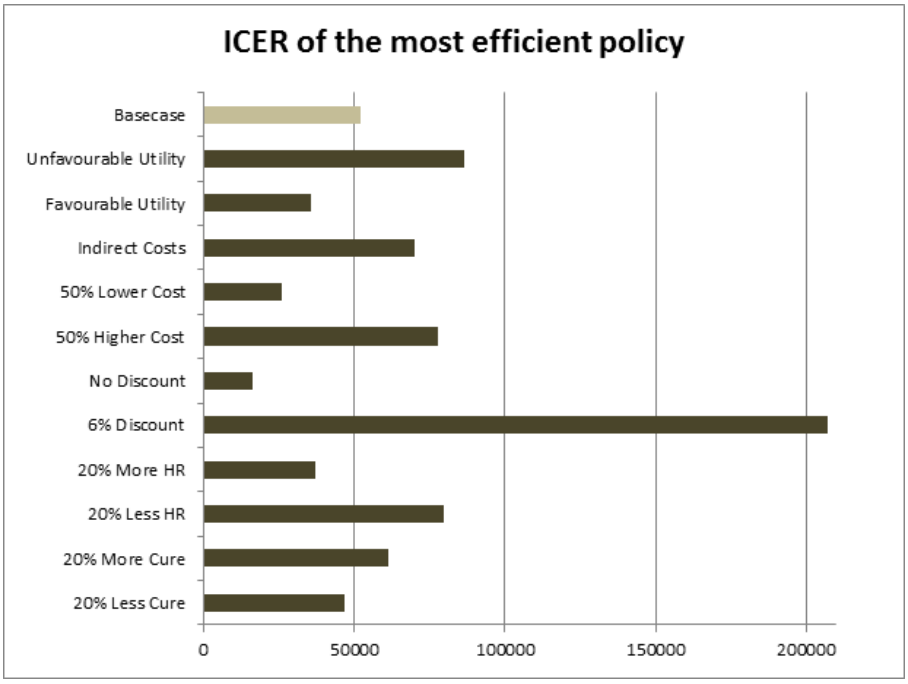
Screen Policy	Treatm. &	Over dx	PCM reduction (%)	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
55-69, 4-year§	AS, 3-year		11	13.8	10.0	0.52	51,645	51,645
	AS, Interm		11	13.6	9.8	0.52	53,304	Domin.
	AS, yearly		12	14.6	10.3	0.54	51,991	64,436
	IRT	11	12	15.2	9.9	0.55	56,171	Domin.
55-61, Yearly	AS, 3-year		9	12.9	10.4	0.83	80,035	Weak D
	AS, Interm		9	13.1	10.5	0.85	80,857	Weak D
	AS, yearly		10	14.0	11.1	0.86	76,869	Weak D
	IRT	6	10	15.4	11.3	0.87	77,500	Weak D
55-69, 2-year	AS, 3-year		17	20.4	14.1	0.94	66,310	104,628
	AS, Interm		16	19.5	13.7	0.95	68,912	Domin.
	AS, yearly		18	21.7	14.5	0.97	67,220	105,695
	IRT	18	19	22.8	13.4	1.02	75,620	Domin.
55-69, Yearly	AS, 3-year		20	24.6	17.2	1.61	93,443	Weak D
	AS, Interm		20	24.1	16.9	1.63	96,816	Domin.
	AS, yearly		22	26.9	17.9	1.66	93,108	203,135
	IRT	23	23	28.7	16.2	1.73	106,822	Domin.

* Effects per 1000 screened men. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN. See Supplement Table 4 for additional effects.
& IRT, denotes immediate radical treatment, which can be either radical prostatectomy or radiation therapy with equal probability. AS denotes Active Surveillance. "yearly" denotes yearly biopsies, "3-year" denotes a biopsy every three years after the first year and "Interm" denotes admitting low and intermediate risk patients for AS, with yearly biopsies.
§ Screening at ages 55, 59, 63 and 67.
# Includes screening costs, treatment costs and palliative therapy costs, which we assume was given to all men who died of prostate cancer. Net costs were calculated by the difference in total costs in a situation without screening and the total costs in a situation with screening.
** "Domin." denotes dominated. A policy is classified as, dominated, if there is another policy that has a lower cost and results in more QALY's gained. "Weak D" denotes weakly dominated policies, that is, they are less effective policies that have a higher cost-effectiveness ratio than the next ranked policy.

Multivariate Sensitivity Analyses

In the multivariate sensitivity analyses we focus on screening between 55 and 69 every 4 years. If the set of "unfavorable utilities" is used, the (incremental) cost per QALY significantly increases in all cases, and AS with yearly biopsies which was efficient in the basecase, becomes dominated. With the set of "favorable utilities" immediate treatment dominates the other alternatives. (Supplementary Information Table 2, Figure 1). Computing the cost-effectiveness with a 6% discount rate resulted in a situation where no policy has a cost per QALY lower than $100,000. If there was no discounting, AS would lose its advantage and immediate treatment would become cost-effective. (Supplementary Information Table 2, Figure 1).

Figure 1: Incremental Costs per QALY gained (ICER) by screening and Active Surveillance under different modelling assumptions.* §



*Basecase includes the cost per QALY of screening between ages 55 and 69, every 4 years, with a PSA threshold for biopsy referral of 4 and a biopsy compliance of 41% (in the screening phase), compared with no screening and immediate treatment. The AS protocol in the basecase consists of yearly biopsies, with biopsy compliance given by observed PRIAS data [21].

We also varied several sets of model parameters. In all situations immediate radical treatment and AS for low and intermediate risk men remain dominated. In particular, a lower hazard ratio for treatment and baseline survival (more lives saved by treatment), results in a higher cost per QALY in all cases. Varying the probabilities of referral from AS to immediate treatment has a low impact on the cost per QALY of AS (Supplementary Information Table 3, Figure 1).

Discussion

In this study we estimate the costs and effectiveness associated with screening followed with active surveillance or immediate treatment, compared with no

screening. We find that AS is more cost-effective than Immediate Treatment. In most cases, AS results in more QALYs and less costs than immediate treatment. We find two screening and AS policies that result in a cost per QALY lower than $100,000. This is the case if we screen every 4 years between ages 55 and 69 and refer low-risk men to AS with yearly or triannual biopsies. Referring intermediate risk men for AS, does not seem to be cost-effective. While the AS protocol with triannual biopsies seems more efficient, the differences between the AS protocols are small.

The gain of AS is explained by the overall lower costs, due to less treatment costs, combined with the fact that it maintains a similar or higher level of effectiveness than immediate radical treatment, when we discount the effects at 3% and adjust life years gained with quality of life loss due to treatment.

Our sensitivity analyses show that these findings are robust to changes in the most important model parameters, utilities and costs, though the most efficient screening and AS protocol may change. The only scenarios where AS would lose its advantage relative to immediate treatment would occur if we would not use discounting or if we would use the set of more favourable utilities towards screening and treatment. Screening every 4-years with AS may result in an incremental cost per QALY higher than $100,000 if we would increase the discount rate to 6%.

Previous literature suggests that AS has the potential to significantly reduce the harms of prostate cancer screening, while keeping a large portion of the benefit [7,9-13]. Some studies focused on the costs of AS relative to other therapies [12, 13, 24, 26-29]. The majority of these studies report that AS has lower costs than immediate treatment, a fact which we also found in our study. Only two studies focused on the cost-effectiveness of AS [12,13].

Our study is the first to link the cost-effectiveness of AS with the screening schedule and to compare the cost-effectiveness of different AS protocols. Furthermore, our model has a natural history of prostate cancer based on a large randomized control trial, with probabilities of progression while on AS, based on the JH cohort. In contrast to [12,13] we model AS explicitly. Additionally [12] was criticized by using a debatable set of assumptions [30].

The results of this study are also subject to some limitations. The costs of treatment were obtained from the lifetime estimates from another microsimulation model [24]. We found the utility estimates have a significant effect in the cost per QALY. In particular, the results are sensitive to the utility of post-recovery of treatment, since it has a duration of 9 years, by assumption.

Our AS model uses a simplification of the actual criteria for selection and later, referral to treatment. For instance, while our model uses PSA, T-stage and Gleason, as selection criteria, most AS cohorts also use number of positive biopsy cores or PSA density [6].

The probabilities of referral to treatment while in AS are based on a single cohort [3]. Other cohorts have slightly different populations, usually with less strict selection criteria than JH, and may have different follow-up protocols [6]. For instance, in the Toronto cohort [4], which does a biopsy every 3 or 4-years, the treatment-free survival is significantly higher than in JH.
Our model of AS assumes 100% referral of low-risk men to AS. The pattern of usage of AS by low-risk men seems to be changing rapidly, as it went from 14% to about 40% in the CaPSURE registry in the period 2010-2013 [31], but it is still far from 100%.

Prostate cancer screening is still a controversial topic. The USPSTF [32] recommended against PSA-based prostate cancer screening, however, in a previous study it was suggested that limited screening could be cost-effective [33]. This study did not model the effect of including all low risk men in AS. Since we found that for most screening policies, referring low-risk men for AS is more cost-effective, adding AS to [33] could change the optimal age to stop screening to a later age.

We believe there is room for improvement in the cost per QALY [34]. For instance, Siddiqui et al [35] found that the so called “targeted” MR-guided biopsy has a significantly higher sensitivity and specificity for intermediate and high-risk tumors than standard biopsy, while diagnosing less low-risk prostate cancers. The application of this technology could therefore substantially reduce the amount of progression during AS.

In this study we found that AS is more cost-effective than immediate treatment, after considering a variety of screening policies and AS protocols. Out of the three AS protocols considered, it seems that offering AS to low-risk men and triannual biopsies is more efficient, however, the differences between protocols are relatively small and may change depending on the screening schedule. Additionally, this study suggests that limited screening (every 4 years) together with AS could be cost-effective at a willingness to pay threshold of $100,000.

References

1. Bul M, Zhu X, Rannikko A, et al (2012) Radical Prostatectomy for Low-Risk Prostate Cancer Following Initial Active Surveillance: Results From a Prospective Observational Study. Eur Urol 62:195-200.

2. Cooperberg M, Cowan J, Hilton J, et al (2011) Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 29:228-34.

3. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

4. Klotz L, Vesprini D, Sethukavalan P, et al (2015) Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 33:272-7.

5. Dall'Era MA, Albertsen PC, Bangma C, et al (2012) Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol 62:976-83.

6. Simpkin AJ, Tilling K, Martin RM, et al (2015) Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 67:993-1005.

7. Cooperberg MR, Carroll PR, and Klotz L. Active Surveillance for Prostate Cancer: Progress and Promise. J Clin Oncol 2011;29:3669-76.

8. Klotz L (2015) Active Surveillance for Prostate Cancer: Debate over the Application, Not the Concept. Eur Urol. 67:1006-8.

9. Xia J, Trock BJ, Cooperberg MR, et al (2012) Prostate cancer mortality following active surveillance versus immediate radical prostatectomy. Clin Cancer Res 18:5471-8.

10. de Carvalho TM, Heijnsdijk EAM and de Koning HJ (2016) Estimating the risks and benefits of Active Surveillance protocols for Prostate Cancer: A microsimulation study. BJU Int. doi: 10.1111/bju.13542.

11. Heijnsdijk EA, Wever EM, Auvinen A, et al (2012) Quality of Life Effects of Prostate Specific Antigen Screening. N Engl J Med 367:595-605.

12. Hayes JH, Ollendorf DA, Pearson SD, et al (2013) Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis. Ann Intern Med. 158:853-60.

13. Roth JA, Gulati R, Gore JL, et al (2016) Economic Analysis of Prostate-Specific Antigen Screening and Selective Treatment Strategies. JAMA Oncol. doi: 10.1001/jamaoncol.2015.6275.

14. de Carvalho TM, Heijnsdijk EAM and de Koning HJ (2016) Estimating the individual benefit of immediate treatment or active surveillance for prostate cancer after screen-detection in older (65+) men. Int J Cancer. 138:2522-8.

15. Draisma G, Postma R, Schröder FH, et al (2006) Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer 119:2366-71.

16. Draisma G, Boer R, Otto SJ, et al (2003) Lead time and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst 95:868-78.

17. de Carvalho TM, Heijnsdijk EAM and de Koning HJ (2015) Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 136:1600-7.

18. Bill-Axelson A, Holmberg L, Garmo H, et al (2014) Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 370:932-42.

19. Etzioni R, Gulati R, Tsodikov A, et al (2012) The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 118:5955-63.

20. Pinsky PF, Andriole GL, Kramer BS, et al (2005) Prostate, Lung, Colorectal and Ovarian Project Team. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 173:746-50.

21. Bokhorst LP, Alberts AR, Rannikko A, et al (2015) PRIAS study group.Compliance Rates with the Prostate Cancer Research International Active Surveillance (PRIAS) Protocol and Disease Reclassification in Noncompliers. Eur Urol. 68:814-21.

22. Sanda MG, Dunn RL, Michalski J, et al (2008) Quality of life and satisfaction with outcome among prostate-cancer survivors.N Engl J Med. 358:1250-61.

23. Stewart ST, Lenert L, Bhatnagar V, et al (2005) Utilities for prostate cancer health states in men aged 60 and older. Med Care. 43:347-55.

24. Laviana AA, Ilg AM, Veruttipong D, et al (2015) Utilizing time-driven activity-based costing to understand the short- and long-term costs of treating localized, low-risk prostate cancer. Cancer doi: 10.1002/cncr.29743.

25. Yabroff KR, Lamont EB, Mariotto A, et al (2008) Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 100:630-41.

26. Koerber F, Waidelich R, Stollenwerk B, et al (2014) The cost-utility of open prostatectomy compared with active surveillance in early localised prostate cancer. BMC Health Serv Res. 14:163.

27. Corcoran AT, Peele PB, Benoit RM (2010) Cost comparison between watchful waiting with active surveillance and active treatment of clinically localized prostate cancer. Urology. 76:703-7.

28. Eldefrawy A, Katkoori D, Abramowitz M, et al (2013) Active surveillance vs. treatment for low-risk prostate cancer: a cost comparison. Urol Oncol. 31:576-80.

29. Keegan KA, Dall'Era MA, Durbin-Johnson B, et al (2012). Active surveillance for prostate cancer compared with immediate treatment: an economic analysis. Cancer. 118:3512-8.

30. Pinsky P (2013) Observation versus initial treatment for prostate cancer. Ann Intern Med. 159:574.

31. Cooperberg MR, Carroll PR (2015) Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA 314:80-2.

32. Moyer VA; U.S. Preventive Services Task Force (2012) Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement Ann Intern Med. 157:120-34.

33. Heijnsdijk EA, de Carvalho TM, Auvinen A, et al (2015) Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. J Natl Cancer Inst.;107:366.

34. Bangma CH, Valdagni R, Carroll PR, et al (2015) Active surveillance for low-risk prostate cancer: developments to date. Eur Urol. 67:646-8.

35. Siddiqui MM, Rais-Bahrami S, Turkbey B, et al (2015) Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. JAMA. 313:390-7.

Supplementary Information

Additional Outcomes

Table 1: Effects of immediate radical treatment (IRT) compared to Active Surveillance (AS), per screening intensity. *

	Treatment &	Dx	Screen-Detected	PCM	LYs	QALYs
55-69, 4-year §	AS, 3-year			29	33	31
	AS, Interm			29	31	29
	AS, yearly			29	35	33
	IRT	149	27	29	36	33
55-61, Yearly	AS, 3-year			29	30	29
	AS, Interm			29	30	29
	AS, yearly			29	33	33
	IRT	142	19	29	35	34
55-69, 2-year	AS, 3-year			27	49	45
	AS, Interm			27	46	43
	AS, yearly			26	52	48
	IRT	154	44	26	55	48
55-69, Yearly	AS, 3-year			26	58	54
	AS, Interm			26	55	50
	AS, yearly			25	64	58
	IRT	158	53	25	68	60

* Effects per 1000 screened men. Effects are shown undiscounted. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.
§ Screening at ages 55, 59, 63 and 67.
& Dx denotes diagnosis, IRT, denotes immediate radical treatment which can be either radical prostatectomy or radiation therapy. AS denotes Active Surveillance. "yearly" denotes yearly biopsies, "3-year" denotes a biopsy every three years after the first year and "Interm" denotes admitting low and intermediate risk patients for AS, with yearly biopsies.

Sensitivity Analyses

Table 2: Costs and Effects of immediate radical treatment (IRT) compared to Active Surveillance (AS), for men screened between ages 55 and 69 every 4 years, under several asssumptions.*

Assumption	Treatment&	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
Favorable Utility	AS, 3-year	14.6	14.6	0.54	36,650	Weak D
	AS, Interm	13.6	13.6	0.55	40,382	Domin
	AS, yearly	15.4	15.3	0.56	36,162	Weak D
	IRT	16.0	15.9	0.57	35,834	35,834
Unfavorable Utility	AS, 3-year	14.6	6.2	0.54	86,460	86,460
	AS, Interm	13.6	5.7	0.55	95,870	Domin
	AS, yearly	15.4	6.1	0.56	90,289	Domin
	IRT	16.0	5.3	0.57	107,360	Domin
50% lower Costs	AS, 3-year	14.6	10.3	0.27	25,959	25,959
	AS, Interm	13.6	9.6	0.27	28,526	Domin
	AS, yearly	15.4	10.6	0.28	26,199	34,690
	IRT	16.0	10.2	0.29	27,836	Domin
50% higher Costs	AS, 3-year	14.6	10.3	0.80	77,877	77,877
	AS, Interm	13.6	9.6	0.82	85,579	Domin
	AS, yearly	15.4	10.6	0.83	78,597	104,070
	IRT	16.0	10.2	0.85	83,509	Domin
Including Indirect costs	AS, Interm	13.6	9.6	0.72	74,949	Weak D
	AS, 3-year	14.6	10.3	0.72	69,995	69,995
	AS, yearly	15.4	10.6	0.74	70,147	75,525
	IRT	16.0	10.2	0.75	73,296	Domin
No discount	AS, 3-year	33.1	30.5	0.51	16,863	Weak D
	IRT	36.4	32.5	0.53	16,451	16,451
	AS, yearly	35.0	31.8	0.54	16,873	Domin
	AS, Interm	31.1	28.3	0.55	19,605	Domin
6% Discount	AS, Interm	6.4	2.4	0.51	208,036	Weak D
	AS, 3-year	6.9	2.5	0.51	207,304	207,304
	AS, yearly	7.3	2.4	0.52	218,317	Domin
	IRT	7.6	1.7	0.55	325,765	Domin

* Effects per 1000 screened men. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.
& IRT, denotes immediate radical treatment, which can be either radical prostatectomy or radiation therapy with equal probability. AS denotes Active Surveillance. "yearly" denotes yearly biopsies, "3-year" denotes a biopsy every three years after the first year and "Interm" denotes admitting low and intermediate risk patients for AS, with yearly biopsies.
# Includes screening costs, the difference in treatment costs between the situation with and without screening and the difference in costs due to the reduction in palliative therapy, which we assume was given to all men who died of prostate cancer.

Table 3: Costs and Effects of immediate radical treatment (IRT) compared to Active Surveillance (AS), for men screened between ages 55 and 69 every 4 years, under several asssumptions.*

Assumption	Treatment &	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
More Cure (20%)	AS, 3-year	15.5	11.4	0.53	46,607	46,607
	AS, Interm	14.5	10.6	0.54	51,129	Domin
	AS, yearly	16.4	11.7	0.55	46,944	58,099
	IRT	17.0	11.3	0.56	49,694	Domin
Less Cure (20%)	AS, 3-year	13.3	8.8	0.54	61,465	61,465
	AS, Interm	12.4	8.2	0.56	67,620	Domin
	AS, yearly	14.0	9.0	0.56	62,184	90,038
	IRT	14.6	8.7	0.58	66,529	Domin
Lower HR Treatment (20%)	AS, 3-year	11.7	6.9	0.55	79,527	79,527
	AS, Interm	10.9	6.5	0.56	86,975	Domin
	AS, yearly	12.3	7.0	0.57	81,319	196,617
	IRT	12.8	6.5	0.59	90,219	Domin
Higher HR Treatment (20%)	AS, 3-year	17.8	14.0	0.52	36,902	36,902
	AS, Interm	16.6	13.0	0.53	40,772	Domin
	AS, yearly	18.8	14.5	0.54	36,963	38,694
	IRT	19.6	14.3	0.55	38,333	Domin
Higher Probabilities of referral	AS, 3-year	14.8	10.4	0.54	51,730	51,730
	AS, Interm	14.0	9.8	0.54	56,437	Domin
	AS, yearly	15.5	10.6	0.56	52,496	87,596
	IRT	16.0	10.2	0.57	55,673	Domin
Lower Probabilities of referral	AS, 3-year	14.4	10.2	0.53	52,172	52,172
	AS, Interm	13.2	9.4	0.55	57,645	Domin
	AS, yearly	15.2	10.6	0.55	52,242	54,185
	IRT	16.0	10.2	0.57	55,673	Domin

* Effects per 1000 screened men. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

& IRT, denotes immediate radical treatment, which can be either radical prostatectomy or radiation therapy with equal probability. AS denotes Active Surveillance. "yearly" denotes yearly biopsies, "3-year" denotes a biopsy every three years after the first year and "Interm" denotes admitting low and intermediate risk patients for AS, with yearly biopsies.

Includes screening costs, the difference in treatment costs between the situation with and without screening and the difference in costs due to the reduction in palliative therapy, which we assume was given to all men who died of prostate cancer.

** ICER stands for, Incremental Cost-Effectiveness Ratio. "Domin" denotes dominated. A policy is classified as "Dominated" if there is another policy that has a lower cost and results in more QALY's gained. "Weak D" denotes weakly dominated policies, that is, they are less effective policies, that have a higher cost-effectiveness ratio than the next ranked policy.

Chapter Nine

When should Active Surveillance for prostate cancer stop if no progression is detected?

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk,
Harry J. de Koning.

Accepted for publication. The Prostate

Abstract

Background: A significant proportion of screen-detected men with prostate cancer may be overdiagnosed. Active Surveillance (AS) has emerged as a way to mitigate this problem, by delaying treatment of men, who are at low-risk until this becomes necessary. However, it is not known after how much time or biopsy rounds should patients stop AS and transition to conservative management (CM), if no progression is detected.

Methods: We used a microsimulation model with natural history of prostate cancer based on ERSPC and SEER data. We modeled referral to treatment while in AS, based on Johns Hopkins treatment-free survival data. We projected lifetime costs and effects of AS (and radical treatment, if progression is detected) under different biopsy follow-up schedules compared to CM, where radical treatment only occurs when men would be clinically diagnosed in absence of screening.

Results: For men with low-risk disease in younger age groups (55-65), AS is cost-effective for up to 7 yearly biopsy rounds. For men older than 65, even one biopsy round results in quality adjusted life years (QALYs) lost, though it may result in QALYs gained for men without previous screening. For men with intermediate risk disease AS is cost-effective even for men in 65-75 age group.

Conclusion: The benefit of AS when compared to CM is strongly dependent on life expectancy and disease risk. Clinicians should take this into account when selecting men to AS, deciding on biopsy frequency and when to stop AS surveillance rounds and transition to CM.

Introduction

Active Surveillance (AS) has emerged as a way to minimize overtreatment due to frequent PSA-based prostate cancer screening. It consists of the monitoring of newly diagnosed and not yet treated men through PSA tests and/or repeat biopsies. The goal of AS is to delay or avoid radical treatment in patients who are unlikely to become symptomatic [1]. In the conservative management (CM) regimen patients are also monitored, although not with invasive procedures like, prostate biopsies, and without curative intent, unless the patient becomes symptomatic. It is similar to watchful waiting, though this term often refers only to older men, or men with major comorbidities [1].

The majority of clinical cohorts following men on AS have a limited follow-up, which is not long enough to establish the long term effects [2-7], however there is an emerging consensus regarding the safety of AS for low-risk men [8-10], and the rate of low-risk men assigned to AS is increasing rapidly [11]. Still there is substantial uncertainty about what is the most optimal way of performing AS, namely about whom to include and the follow-up protocol [8-10].

Computer modelling has been used to make projections of the potential effects of AS on prostate cancer mortality (PCM) and overtreatment. Previous studies [12, 13] found that compared to immediate treatment, AS results in a modest increase in PCM (1.4% and 1.8%, respectively). Others [14-17] have compared the costs or cost-effectiveness of AS and/or CM to immediate treatment and found that AS seems to be more advantageous. As far as we know, no simulation study has tried to optimize the age to stop AS and start CM.

There are virtually no clinical studies on when can a patient safely discontinue AS, if no progression is detected and stop being considered for radical treatment, which may potentially lead to treatment decisions to depend on personal or physician's preferences. In the PRIAS cohort, the compliance with PSA testing during AS was relatively high, however the compliance with prostate biopsies is relatively low and decreases rapidly over time from 81% at 1 year to 33% after 10 years [18]. A recent study [19] using SEER data found that only 13% of men underwent prostate biopsy after 2 years, in a community setting. For older patients or men suffering from significant comorbidities, this is probably reasonable, however, for patients with a longer life expectancy there could be a danger of progression to advanced disease, due to non-adherence to the biopsy protocol.

The aim of this study is to determine at which age it is safe and more cost-effective to leave AS and transition to a CM context, for different age and disease risk groups. We use a previously validated [12, 20] microsimulation model of prostate

cancer screening and AS, and compare the incremental cost-effectiveness of each additional AS biopsy round with CM.

Methods

Simulation Model

MISCAN is a microsimulation model designed to evaluate the lifetime effects of prostate cancer screening. A detailed description is available in the Model Appendix, in http://cisnet.cancer.gov/prostate/profiles.html and in previously published studies [12, 21].

Men may experience onset of prostate cancer, based on a constant hazard per age group. In each disease stage, a man may progress to a higher T-stage, a higher grade, become metastasized or clinical diagnosed. We modeled 18 health states corresponding to the combination of 3 stages (T1, T2 and T3), 3 grades (Gleason less than 7, 7, and more than 7) and whether or not cancer is metastasized.

Additional states were created specifically to model AS. Men in T2-stage, Gleason 6 were classified as T2a or T2bc and Gleason 7 men as classified as Gleason 3+4 or 4+3, depending on their remaining lead-time (i.e. time to clinical diagnosis in absence of screening) and age group, based on ERSPC data. We also modeled PSA growth from onset until detection. The parameters of the natural history, which include onset of the disease, durations, transition probabilities between health states and PSA growth were calibrated to observed ERSPC's PSA distribution data and SEER incidence data [21, 22] and Model Appendix Figures 2 and 3.

At clinical detection, a baseline survival is assigned, which depends on Gleason score (< 8 or ≥ 8) and was estimated based on SEER data from the pre-PSA era (1983-1986). This was adjusted, for improvements on survival not directly associated with screening or primary treatment, by adding a hazard ratio for prostate cancer survival of 0.82, which was calibrated to the observed PCM in the ERSPC control (no screening) group (Model Appendix Table 5-7).

We add a benefit of treatment and a benefit of screening (in case the patient is screen-detected) to the baseline survival. The hazard ratio for the benefit of treatment equals 0.56 for radical prostatectomy [23]. For radiation therapy this equals 0.63 using the same rationale as in [24]. The effect of screening is modelled through a lead-time dependent cure rate for non-metastatic cases,
Cure probability = exp(cure parameter * lead-time).

The cure parameter was calibrated based on the observed prostate cancer mortality reduction due to screening in the ERSPC trial [25] (Model Appendix Tables 5-7 and Model Appendix Figure 4).

When the patient is referred to AS, natural history progresses as if the patient was not screen-detected. Referral to radical treatment may occur due to detection of Gleason or volume progression (which is assumed to occur after an increase in stage) at each biopsy round, due to personal preference (randomly selected from all men in AS) or due to clinical detection at the time (Model Appendix Table 8). The probabilities of referral to radical treatment for low-risk men are estimated based on JH-AS observed treatment-free survival data [4] (Model Appendix Table 9). For intermediate risk men, we assume that given the risk of progression (which is Gleason and T-stage dependent), the probabilities of referral to treatment are the same as for low-risk men.

At the time of referral to treatment, survival differs compared to what would have happened if the patient would be treated at the time of screening diagnosis, by assuming a decrease in the benefit of screening, depending on the remaining lead-time for the patient.

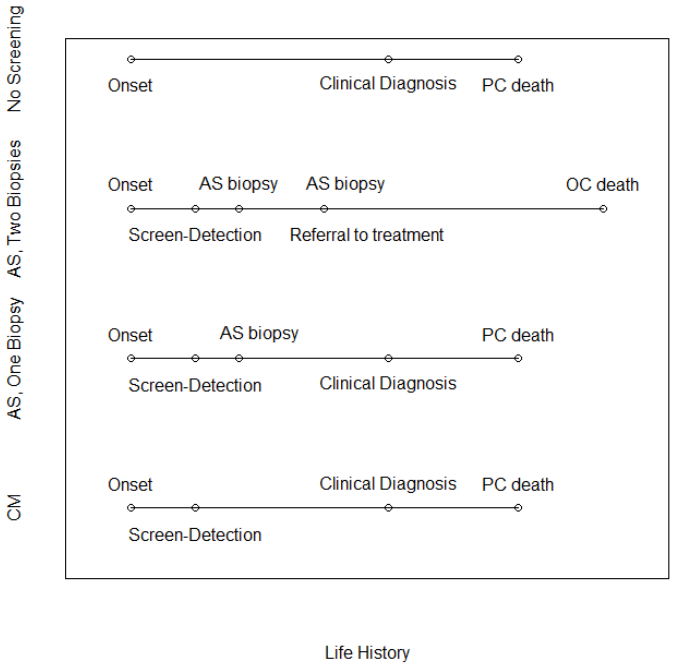
Active Surveillance

We simulate a cohort of 10 million men, based on US lifetables. In the basecase, men are screened between 55 and the upper bound of the age group selected for AS (with an attendance rate of 90%), with a PSA threshold for biopsy referral of 4 and biopsy compliance equal to 41%, based on the PLCO trial [26]. Biopsy Compliance during AS is based on the PRIAS study [17] and Model Appendix Table 11.
We only follow men who are selected for AS in a particular age group (55-59, 60-64, 65-69, 70-74). Men selected for AS have either low risk (Gleason 6, <= T2 stage, PSA <10) or intermediate risk (up to Gleason 3+4, T2 stage and PSA < 20) disease. We compare AS protocols differing by follow-up (1, 4, 7 and 10 biopsy rounds) and biopsy frequency (annual or triannual) with CM, where men are only treated when clinical diagnosis would occur in absence of screening (See Figure 1, for an example).

Initially, the rate of clinical diagnosis in absence of screening was based on the control group and interval cancers in the screening group of the ERSPC trial. Since the model was adapted to US, an extra hazard of clinical diagnosis was added, reflecting an earlier probability of detection. This was calibrated to SEER incidence data from before the introduction of PSA-based screening.

We assume that the main difference between AS and CM surveillance regimens is that AS includes prostate biopsies, i.e. we assume a similar rate of PSA tests or visits to the doctor between the two regimens. We also assume that the main driver of referral to treatment while in AS is the result of the prostate biopsy.

Figure 1: An example of what would happen to a man who would experience clinical diagnosis and prostate cancer death in absence of screening, by follow-up protocol.



* In "AS, One Biopsy" protocol, AS stops after one biopsy, in "AS, two biopsies" protocol, AS stops after two biopsies.
& In this example we show the life history of a men who would be clinically diagnosed in absence of screening, and die from prostate cancer, but that could be saved by treating early. This men also experiences progression during AS, after the time of the first biopsy.

Quality of Life and Costs

QALYs were calculated by using utility estimates ranging between 0 (death or worst imaginable health) and 1 (pre-diagnosis health, which is assumed to be full health). Estimates for the utilities and its durations for all stages of AS, including, prostate biopsies, radical treatment, post-recovery and palliative therapy are based on [27-29] (Table 1). QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

Table 1: Utilities, durations and costs of screening and treatment

Event	Utility	Duration (Years)	Cost (US dollars)
Biopsy	0.90	0.06	743
RT , < 2 months	0.73	0.16	23,565#
RT , 2-12 months	0.78	0.84	
RP , < 2months	0.67	0.16	16,946#
RP, 2-12 months	0.77	0.84	
AS (Surveillance Costs)§	-	6	245 (per year)
Post-Recovery	0.95	9	
Palliative Therapy	0.60	2.5	48,472 &
Terminal Illness	0.40	0.5	

* All utilities and durations based on [27-29].
# Cost of radical prostatectomy (RP) and radiation therapy (RT) includes surveillance costs [15].
§ Surveillance costs of AS, include 4 PSA tests and one visit to the doctor per year, for 6 years.
& Based on Yabroff *et al* [31].

Costs of AS compared to CM include the cost of prostate biopsies and the extra cost of radical treatment, since in CM regimen only men who are clinically diagnosed are treated. The cost of immediate treatment is the average lifetime cost for men treated at age 65, including adverse events, indirect costs and post-treatment surveillance [30]. Costs of palliative therapy are based on [31]. All costs were inflated to 2015 US Dollars.

We calculate the cost per QALY using a 3% discount rate for all costs and effects. The most efficient AS policies are determined based on their incremental cost-effectiveness ratio (ICER). An AS protocol is considered to be cost-effective if its ICER is below $100,000.

Outcomes

We estimated the cost-effectiveness of each AS protocol compared to CM (no treatment) for men who are selected to AS. The main outcome is cost per QALY gained of the AS protocol compared to CM. Other important outcomes include overtreatment, prostate cancer mortality, life years gained (LYG) and costs.

Sensitivity Analyses

We analyzed the effects of some plausible alternative scenarios. For the age groups 55-59 and 60-64, we studied whether a lower biopsy frequency (triannual after the first year) or a higher biopsy compliance (81% for the whole follow-up) would have

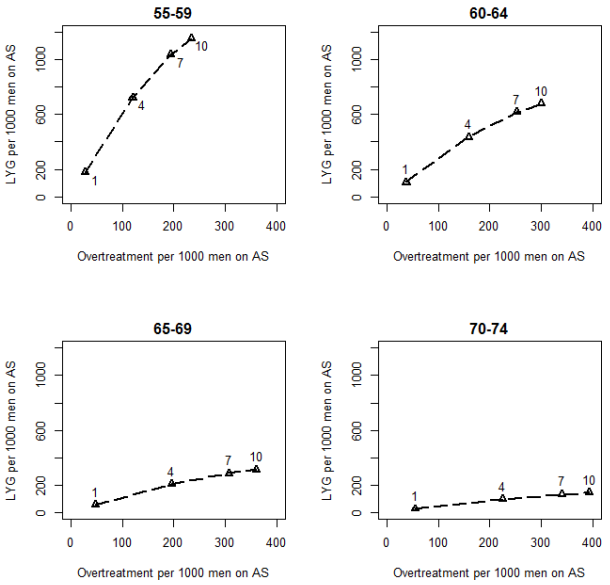
a large impact on QALYs gained. For the age group 65-69, we projected the effects of starting screening at age 65, and we split the age group in smaller intervals (65-66, 65-67, 65-68). In order to assess the effect of uncertainty around the parameter estimates on the outcomes, several multivariate sensitivity analyses were performed, including the utility and cost estimates for each event, the parameters of the model related to treatment benefit, the benefit due to early detection, and the rate of clinical diagnosis (Supplemental Information Table 1).

Results

Effects

In figure 2, LYG compared to the CM situation are plotted against overtreated men per 1000. Both outcomes are highly dependent on the age group of selection for AS. In the age group 55-59, performing 4 biopsy rounds results in 723 LYG at a cost of 120 overtreated men per 1000 men in AS. By contrast, for the age group 70-74, 4 biopsy rounds result in only 98 LYG with 224 overtreated men per 1000.

Figure 2: Life Years Gained (LYG) and Overtreatment for Low-Risk Men (≤ T2a Gleason 6, PSA <10) in AS by number of yearly biopsy rounds (1, 4, 7 and 10 rounds) compared with CM.



* All values are undiscounted. Biopsy frequency is yearly. The points in the figure are respectively after 1, 4, 7 and 10 biopsy rounds.

In table 2, all effects are shown, with QALYs and LYG discounted at 3%. AS results in life years gained for all age groups. However after adjusting life years for quality life, our model projects that for men older than 65, even one biopsy round results in QALYs lost, relative to CM situation. For men aged between 55-59 after 4 biopsy rounds, the model projects about 264 QALYS gained and for men aged between 60-64, 118 QALYs gained.

Table 2: Costs and Effects of Active Surveillance (AS) for low-risk men per number of biopsy rounds.*

Age Group	Biopsy Rounds	Treat. &	Overtreat.	PCM (%)	LY	QALY	Total Cost#	Mean Cost per QALY	ICER **
55-59	1	626	28	16	81	64	0.9	13,755	WD
	4	719	120	11	326	264	2.8	10,665	10,665
	7	793	193	8	461	368	4.0	10,875	11,412
	10	833	233	6	511	402	4.6	11,331	16,199
60-64	1	509	38	11	52	27	1.1	41,318	WD
	4	631	159	7	209	118	3.7	31,684	31,684
	7	724	252	5	291	157	5.3	33,927	40,658
	10	773	300	4	319	163	6.1	37,106	113,909
65-69	1	399	48	7	29	-3	1.4	-	-
	4	548	196	4	105	-14	4.6	-	-
	7	658	306	3	144	-31	6.6	-	-
	10	711	359	2	157	-43	7.5	-	-
70-74	1	310	55	4	15	-22	1.5	-	-
	4	479	224	3	54	-84	5.2	-	-
	7	594	339	2	73	-124	7.3	-	-
	10	647	392	2	79	-141	8.2	-	-

* Effects per 1000 screened men (unless denoted otherwise). Costs and Effects (LYs and QALYs) are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.
# Includes the cost of prostate biopsy, treatment costs and palliative therapy costs, which we assume was given to all men who died of prostate cancer. Net costs were calculated by the difference in total costs in the AS and CM situation.
** "W D" denotes a weakly dominated strategy.
& Includes men who are clinically diagnosed, during and after AS follow-up.

Costs

In table 2, the total cost of AS is shown. Cost of AS increases with number of biopsy rounds and age group. Cost varied between $0.9 million, per 1000 men in AS, for one biopsy for 55-59 men and $8.2 million for men aged 70-74 after 10 biopsy rounds.

Cost-Effectiveness

We show the cost-effectiveness of AS compared to CM in table 2. For low-risk men in age groups 65-69 and 70-74 AS is not effective. For men in age group 55-59, the estimated ICER after 10 yearly biopsy rounds (yearly biopsies) is $16,199. For the age group 60-64 the ICER is $113,909.

Other Scenarios

Under the assumption that biopsy compliance during AS remains 81% for the whole follow-up (observed biopsy compliance in the first year of the PRIAS [14]), we see that after 4 biopsy rounds QALYs gained increase from 264 to 300 for the 55-59 age group, and from 118 to 131 for the 60-64 age group (Table 3). If the biopsy frequency becomes triannual after the first year, QALYs gained decrease from 264 to 196 for the 55-59 age group, and from 118 to 71 for the 60-64 age group (Table 3). For men aged between 70-74, with intermediate-risk prostate cancer, AS may be cost-effective up to 7 biopsy rounds. For men aged 65-69 the ICER is only $21,299 after 10 biopsy rounds. AS for low-risk men aged 65-69 may results in QALYs gained if screening only started at age 65 or if AS is restricted for men younger than 68 (Supplemental Information Table 2).

Table 3: Costs and Effects of Active Surveillance (AS) per number of biopsy rounds under different scenarios.* &

Age Group and Scenario	Biopsy Rounds	Treated&	Overtreatment	PCM (%)	LY	QALY	Total Cost#	Mean Cost per QALY	ICER**
55-59	4	733	134	10	369	300	3.1	10,451	10,451
Compli.	7	823	224	6	515	406	4.7	11,554	14,688
60-64	4	650	178	7	233	131	4.2	31,709	31,709
Compli.	7	763	291	4	324	168	6.2	36,779	54,936
55-59	2	666	68	14	196	163	1.5	9,239	WD
Triannual	3	706	107	12	275	227	2.0	8,272	8,272
60-64	2	562	92	9	122	71	2.1	29,387	WD
Triannual	3	612	141	8	172	100	2.8	28,103	28,103
65-69	1	634	25	17	53	40	0.7	18,119	18,119
Intermed.	4	720	110	13	211	167	2.2	13,273	13,273
Risk	7	789	178	11	297	230	3.1	13,348	13,545
	10	821	210	10	325	248	3.4	13,905	21,299
70-74	1	539	31	12	31	13	0.9	67,734	WD
Intermed.	4	642	133	10	121	55	2.9	52,145	52,145
Risk	7	719	209	8	167	72	4.0	56,149	69,588
	10	754	244	8	182	74	4.5	60,373	173,882

* Effects per 1000 screened men. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

& "Compli." denotes a scenario where biopsy compliance remains at 81% instead of decreasing through time as observed in [18]. "Triannual" denotes a scenario where biopsies occur every three years after the first year in AS. "Intermediate Risk" denotes men who are Gleason 3+4 and/or have a PSA at diagnosis between 10 and 20.

Includes the cost of prostate biopsy, treatment costs and palliative therapy costs, which we assume was given to all men who died of prostate cancer. Net costs were calculated by the difference in total costs in the AS and CM situation.

** "W D" denotes a weakly dominated strategy.

& Includes men who are clinically diagnosed, during and after AS follow-up.

Sensitivity Analyses

Our sensitivity analyses focus on the age group 60-65. At four biopsy rounds we find a range for the ICER between $15,842 (50% higher costs) and $78,709 (unfavorable utilities for prostate biopsies and treatment). At seven biopsy rounds the ICER ranges between $20,329 and $217,517. Of all the sets of parameters considered, utilities have the largest effect on the cost per QALY (Supplemental Information Table 7 and 8). The only scenario where AS after 7 biopsy rounds was not cost-effective was when we applied unfavorable utilities towards screening and treatment procedures. By contrast, if we would apply more favorable utilities, lower costs, lower treatment benefit and a higher rate of clinical diagnosis then AS would become cost-effective after 10 biopsy rounds (Supplemental Information Table 3 and 4).

Discussion

In this study we determine for the first time, how much time should men stay on AS and be considered for treatment. Previous clinical studies have ascertained AS is relatively safe for low-risk men [2-7, 11, 12]. Additionally, most studies [14-17] found that AS is less expensive or more cost-effective than immediate treatment. However, no clinical cohort or simulation study has examined how much time should men be on AS or how the age group or life expectancy could determine the intensity of the follow-up schedule.

In this study we find that AS with 7 yearly biopsy rounds for low-risk men is cost-effective compared to CM, but only for age group 55-65 and for men up to age 75 with intermediate risk disease. We performed this by calculating the cost-effectiveness of AS with different follow-up biopsy schedules (1, 4, 7 and 10 yearly biopsy rounds) compared to a CM group where men can only be treated if they would be clinically diagnosed in absence of screening.

The inclusion of low-risk men younger than 60 in AS is not consensual [10]. These are the best candidates for immediate treatment and have a relatively low probability of being overtreated [32, 33]. We find that if these men are included on AS, they require an intensive yearly biopsy schedule, which should continue for at least 10 yearly biopsy rounds. A low biopsy compliance or biopsy frequency may result in QALYs lost and in a significant increase in PCM. By contrast, for men older than 70, AS is not cost-effective, and does not result in QALY's gained, though some lives are saved. For men aged between 60-64, AS appears to be cost-effective up to 7 biopsy rounds, and for men aged 65-69 AS results on QALYs lost. With a finer analysis, we find that for men aged 65-67, and for men who started screening after age 65 AS

results in QALYs gained but it is still not cost-effective. Just like for men younger than 60, selecting intermediate risk men for AS is still under debate [34]. We find that if these men are selected, an intensive AS regime with many follow-up biopsies is recommended (for men younger than 70, at least 10 yearly biopsy rounds).

Sensitivity Analyses found that the results in this study are the most sensitive to the utilities. For instance, using more unfavorable utilities about biopsies and treatment made the ICER become higher than $100,000 at 7 years for the age group 60-65. However, there were more scenarios (higher cure rate, lower treatment effect, favorable utilities), where AS with 10 yearly biopsy rounds became cost-effective. This study is subject to some limitations. Our AS model uses a simplification of the criteria of selection for AS and to be referred to treatment compared to most clinical cohorts. For instance, we do not model the number of positive biopsy cores and we assume volume progression can only occur, if there is an increase in T-stage. The probabilities of referral to treatment are based on the JH cohort, which has slightly different selection criteria to AS and referral to treatment criteria than other cohorts like PRIAS or the Toronto cohort [7]. We also assume that all men whose progression is detected are referred to radical treatment, which may not happen in clinical practice [4].

Changes in the rate of clinical diagnosis could affect the cost-effectiveness of the intervention. We examined the effect of parametric uncertainty around the hazard of clinical diagnosis parameters in predicted QALYs and costs. We found that while the proportion of men treated may change significantly (Supplemental Information Table 8) the cost per QALY did not vary substantially when we changed the values of the clinical diagnosis parameters by 20%.

These results apply for the average US population. Populations at higher risk, for instance African-Americans, may need a more intensive schedule compared to their Caucasian peers, for the same age and disease group [10, 35]. Comorbidity level should also be taken into account when building a personalized AS biopsy schedule. While most clinical cohorts find that AS for low-risk men is safe, there is still some debate about the most optimal way to perform AS. In particular, it is not known, when should a person stop AS, if no progression is detected.

We conclude that the AS protocol for younger low-risk men (55-65) and for older (65-75) intermediate-risk men could consist of at least 7 yearly biopsy rounds. For older (65-75) low-risk men, AS is not effective compared to CM, after taking into account quality of life outcomes. Therefore for these men, there is no need for an intensive yearly biopsy schedule or to remain for many years on AS.

References

1. 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. 2016. pii: S0302-2838(16)30470-5.

2. Bul M, Zhu X, Rannikko A, et al. Radical Prostatectomy for Low-Risk Prostate Cancer Following Initial Active Surveillance: Results From a Prospective Observational Study. Eur Urol 2012;62:195-200.

3. Cooperberg M, Cowan J, Hilton J, et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. J Clin Oncol 2011;29:228-34.

4. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

5. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. J Clin Oncol 2015;33:272-7.

6. Dall'Era MA, Albertsen PC, Bangma C, et al. Active surveillance for prostate cancer: a systematic review of the literature. Eur Urol 2012;62:976-83.

7. Simpkin AJ, Tilling K, Martin RM, et al. Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 2015;67:993-1005.

8. Klotz L. Active Surveillance for Prostate Cancer: Debate over the Application, Not the Concept. Eur Urol. 2015;67:1006-8.

9. Bangma CH, Valdagni R, Carroll PR, et al. Active surveillance for low-risk prostate cancer: developments to date. Eur Urol. 2015;67:646-8.

10. Carter HB. Optimizing Active Surveillance. Eur Urol 2016. doi: 0.1016/j.eururo.2016.07.017.

11. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA 2015;314:80-2.

12. Xia J, Trock BJ, Cooperberg MR, et al. Prostate cancer mortality following active surveillance versus immediate radical prostatectomy. Clin Cancer Res 2012;18:5471-8.

13. 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. 2016.

14. Roth JA, Gulati R, Gore JL, et al. Economic Analysis of Prostate-Specific Antigen Screening and Selective Treatment Strategies. JAMA Oncol. 2016. doi: 10.1001/jamaoncol.2015.6275.

15. Laviana AA, Ilg AM, Veruttipong D, et al. Utilizing time-driven activity-based costing to understand the short- and long-term costs of treating localized, low-risk prostate cancer. Cancer. 2016;122:447-55.

16. Keegan KA, Dall'Era MA, Durbin-Johnson B, et al. Active surveillance for prostate cancer compared with immediate treatment: an economic analysis. Cancer. 2012;118:3512-8.

17. Koerber F, Waidelich R, Stollenwerk B, et al. The cost-utility of open prostatectomy compared with active surveillance in early localised prostate cancer. BMC Health Serv Res. 2014;14:163.

18. 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.

19. Loeb S, Walter D, Curnyn C, et al. How Active is Active Surveillance? Intensity of Followup during Active Surveillance for Prostate Cancer in the United States. J Urol. 2016;196:721-6.

20. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the individual benefit of immediate treatment or active surveillance for prostate cancer after screen-detection in older (65+) men. Int J Cancer. 2016;138:2522-8.

21. de Carvalho TM, Heijnsdijk EAM and de Koning HJ. Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136:1600-7.

22. Draisma G, Postma R, Schröder FH, et al. Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer 2006; 119:2366-71.

23. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.

24. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012;118:5955-63.

25. 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.

26. Pinsky PF, Andriole GL, Kramer BS, et al; PLCO Project Team. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 2005;173:746-50.

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

28. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med. 2008;358:1250-61.

29. Stewart ST, Lenert L, Bhatnagar V, et al. Utilities for prostate cancer health states in men aged 60 and older. Med Care. 2005;43:347-55.

30. Laviana AA, Ilg AM, Veruttipong D, et al. Utilizing time-driven activity-based costing to understand the short- and long-term costs of treating localized, low-risk prostate cancer. Cancer. 2016;122:447-55.

31. Yabroff KR, Lamont EB, Mariotto A, et al. Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 2008;100:630-41.

32. Gulati R, Inoue LY, Gore JL, et al. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst. 2014;106(2):djt367. doi: 10.1093/jnci/djt367.

33. Wever EM, Heijnsdijk EA, Draisma G, et al. Treatment of local-regional prostate cancer detected by PSA screening: benefits and harms according to prognostic factors. Br J Cancer. 2013;108:1971-7.

34. Musunuru HB, Yamamoto T, Klotz L, et al. Active Surveillance in Intermediate-Risk Patients: Survival Outcomes in the Sunnybrook Experience. J Urol. 2016. pii: S0022-5347(16)31142-9. doi: 10.1016/j.juro.2016.06.102.

35. Maurice MJ, Sundi D, Schaeffer EM, Abouassaly R. Risk of Pathological Upgrading and Upstaging among Men with Low-Risk Prostate Cancer Varies By Race: Results from the National Cancer Data Base. J Urol. 2016. pii: S0022-5347(16)31184-3. doi: 10.1016/j.juro.2016.08.09.

Supplementary Information

Uncertainty included in Sensitivity Analyses

Table 1: Overview of included uncertainty in the multivariate sensitivity analyses*

Parameter	Value	Range
Cure Parameter		
cure parameter	-0.22	Min=-20%, Max = +20%
Hazard Ratio Treatment		
hazard ratio of RP	0.56	0.41, 0.77 #
hazard ratio of RT	0.63	0.46, 0.87 #
Utilities (Favourable , Unfavourable) &		
screening	0.99	1.00, 0.98
biopsy	0.90	0.87, 0.94
cancer diagnosis	0.80	0.85, 0.75
RT , < 2 months	0.73	0.75, 0.71
RT , 2-12 months	0.78	0.88, 0.68
RP , < 2months	0.67	0.78, 0.56
RP, 2-12 months	0.77	0.84, 0.70
post-recovery	0.95	0.93, 0.97
palliative therapy	0.60	0.24, 0.86
terminal illness	0.40	0.24, 0.56

* In an additional separate analysis, all costs are varied by more and less 50%. RP denotes radical prostatectomy, RT denotes radiation therapy.
# Confidence interval for hazard ratio of RP, is the observed in Bill-Axelsson et al (23). The confidence interval of RT was extrapolated using the same ratio as in Etzioni et al (24).
& Adapted from Heijnsdijk et al (27). "Favorable/Unfavorable" refers to whether this utility gives a higher/lower QALY gained by screening, respectively.

Analyses cost-effectiveness age group 65-70

Table 2: Costs and Effects AS by number of biopsy rounds, for men between ages 65 and 69, under several asssumptions.*

Scenario	AS schedule	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
65-66	1 round	34	3	1.3	466,404	WD
	4 rounds	130	13	4.6	340,012	340,012
65-67	1 round	32	0	1.4	46,041,175	WD
	4 rounds	120	2	4.6	2,224,253	2,224,253
65-68	1 round	30	-2	1.4	-	-
	4 rounds	113	-5	4.6	-	-
No Screening	1 round	33	5	1.4	279,439	279,439
Before age 65	4 rounds	115	-5	4.6	-	Domin.

* Effects per 1000 men on AS. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

Sensitivity Analyses

Table 3: Costs and Effects AS by number of biopsy rounds, for men between ages 60 and 64, under several asssumptions.*

Assumption	AS schedule	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
Favorable Utility	1 round	52	47	1.1	23,736	W D
	4 rounds	209	196	3.7	18,964	18,964
	7 rounds	291	271	5.3	19,621	21,349
	10 rounds	319	293	6.1	20,683	33,890
Unfavorable Utility	1 round	52	9	1.1	121,066	W D
	4 rounds	209	47	3.7	78,709	78,709
	7 rounds	291	55	5.3	97,309	217,517
	10 rounds	319	47	6.1	128,411	Domin.
50% lower Costs	1 round	52	27	0.6	20,659	W D.
	4 rounds	209	118	1.9	15,842	15,842
	7 rounds	291	157	2.7	16,963	20,329
	10 rounds	319	163	3.0	18,553	56,955
50% higher Costs	1 round	52	27	1.6	61,977	W D
	4 rounds	209	118	5.6	47,526	47,526
	7 rounds	291	157	8.0	50,890	60,988
	10 rounds	319	163	9.1	55,659	170,863

* Effects per 1000 men on AS. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

Includes the difference in treatment costs between the situation with AS and CM, prostate biopsies and the difference in costs due to the reduction in palliative therapy for men in AS, which we assume was given to all men who died of prostate cancer.

** ICER stands for, Incremental Cost-Effectiveness Ratio. A policy is classified as "Domin." if there is another policy that has a lower cost and results in more QALY's gained. "W D" denotes weakly dominated policies, that is, they are less effective policies, that have a higher cost-effectiveness ratio than the next ranked policy.

Table 4: Costs and Effects AS by number of biopsy rounds, for men between ages 60 and 64, under several asssumptions.*

Assumption	AS schedule	LYs	QALY's	Total Cost#	Average Cost per QALY	ICER **
More Cure (20%)	1 round	54	30	1.1	36,537	W D
	4 rounds	219	129	3.7	28,571	W D
	7 rounds	323	194	5.1	26,426	26,426
	10 rounds	334	180	6.0	33,108	Domin.
Less Cure (20%)	1 round	49	24	1.1	47,822	W D.
	4 rounds	195	101	3.8	37,563	37,563
	7 rounds	272	134	5.4	40,409	49,125
	10 rounds	299	139	6.2	44,408	157,407
Lower HR treatment	1 round	47	22	1.1	52,025	W D
	4 rounds	194	100	3.8	38,091	38,091
	7 rounds	272	133	5.4	40,866	49,297
	10 rounds	298	138	6.2	44,902	158,320
Higher HR treatment	1 round	57	28	1.3	46,473	W D
	4 rounds	228	134	3.8	28,469	28,469
	7 rounds	317	181	5.4	29,559	32,649
	10 rounds	347	190	6.1	32,037	84,142
Higher Rate Clinical Diagnosis	1 round	51	31	1.0	31,421	W D
	4 rounds	197	125	3.1	24,974	24,974
	7 rounds	273	170	4.3	25,312	26,247
	10 rounds	298	180	4.8	26,861	53,324
Lower Rate Clinical Diagnsis	1 round	54	27	1.2	43,065	W D.
	4 rounds	211	109	4.1	37,199	37,199
	7 rounds	295	143	5.9	40,919	52,725
	10 rounds	324	147	6.7	45,730	217,470

* Effects per 1000 men on AS. Costs and Effects are discounted at 3%. Costs are shown in millions of 2015 US dollars. QALYs were calculated by multiplying the loss in utility with the duration of the phase and the number of men who experienced the event, as predicted by MISCAN.

Includes the difference in treatment costs between the situation with AS and CM, prostate biopsies and the difference in costs due to the reduction in palliative therapy for men in AS, which we assume was given to all men who died of prostate cancer.

** ICER stands for, Incremental Cost-Effectiveness Ratio. A policy is classified as "Domin." if there is another policy that has a lower cost and results in more QALY's gained. "W D" denotes weakly dominated policies, that is, they are less effective policies, that have a higher cost-effectiveness ratio than the next ranked policy.

Chapter Ten

Evaluating parameter uncertainty in a simulation model of cancer using emulators

Tiago M. de Carvalho, Eveline A.M. Heijnsdijk, Luc Coffeng, Harry J. de Koning.

Submitted

Abstract

Background: Microsimulation models have been extensively used in the field of cancer modelling. However, there is substantial uncertainty regarding estimates from these models, for example, overdiagnosis in prostate cancer. This is usually not thoroughly examined due to the high computational effort required.

Objective: To quantify the effect of parameter uncertainty on model outcomes, using a computationally efficient emulator (Gaussian Process Regression) instead of the model.

Methods: We use a microsimulation model of prostate cancer (MISCAN) to simulate individual life histories. We analyze the effect of parametric uncertainty on overdiagnosis with probabilistic sensitivity analyses (ProbSA). To minimize the number of MISCAN runs needed for ProbSA, we emulate MISCAN, using data pairs of parameters values and outcomes to fit a Gaussian Process regression model. We evaluate to what extent the emulator accurately reproduces MISCAN by computing its prediction error.

Results: Using an emulator instead of MISCAN, we may reduce the computation time necessary to run a ProbSA by more than 85%. The average relative prediction error of the emulator for overdiagnosis equaled 1.7%. We predicted that 42% of screen-detected men are overdiagnosed, with an associated empirical confidence interval between 38%-48%. Sensitivity analyses show that the accuracy of the emulator is sensitive to which model parameters are included in the training runs.

Conclusions: Even for a model with a large number of parameters and expensive to run, we show that it is possible to conduct a ProbSA including all parameters, within a reasonable computation time by using a Gaussian process regression emulator.

Introduction

Microsimulation Models (MSMs) can be used to describe complex disease processes at the individual patient level. They combine different data sources to project population-level health effects of a novel treatment or intervention compared to standard care. In the field of cancer modelling they have been extensively used to model colorectal cancer, breast cancer and prostate cancer, among others [1-6].

Projections resulting from these models are used to inform health policy decisions, for example regarding early detection recommendations from the USPSTF 7) or updating guidelines from medical associations [8]. However, there is substantial uncertainty regarding estimates from these models, which is usually not thoroughly examined and reported since simulation models of cancer tend to be computationally intensive with a large number of model parameters [1-6].

Usually we distinguish between three types of uncertainty. First-order uncertainty is related to simulation error, and can be eliminated by simulating a large number of disease histories until the effect of individual random draws on the outcomes becomes negligible. In this study, we focus on the effect of parametric or second-order uncertainty in the outcomes of a cancer microsimulation model, that is, we quantify the effect of uncertainty around model parameters on the model outcomes by carrying out a probabilistic sensitivity analysis (ProbSA. The third type, structural or model uncertainty, i.e. uncertainty due to the assumptions used to build the model, can be examined by comparing results across several models [2-4], or by showing a range of results obtained when imposing different sets of assumptions [9].

An example of a substantially uncertain model outcome is overdiagnosis of prostate cancer. A person is overdiagnosed if he would not be clinically diagnosed with cancer in absence of screening. MSMs estimates for overdiagnosis of prostate cancer range between 23-42% of screen-detected cases [10]. Although several study features can affect estimates of overdiagnosis [11], e.g. definition of overdiagnosis, method of estimation or study population, the role of parametric uncertainty is notoriously absent from this debate.

In this study we carry out a ProbSA to quantify the uncertainty due to model parameters in two important outcomes of our simulation model (MISCAN), overdiagnosis and prostate cancer mortality. Since the ProbSA procedure is computationally expensive, we emulate MISCAN using Gaussian Process (GP) regression [12-14] in order to minimize the amount of MISCAN runs. Furthermore, we investigate under which conditions a GP emulator produces reliable estimates of the behavior of a microsimulation model.

Methods

Simulation Model

MIcrosimulation SCreening ANalysis (MISCAN) is a microsimulation model designed to study the effect of screening on incidence and prostate cancer mortality. A detailed description is available elsewhere [6,15], at http://cisnet.cancer.gov/prostate/profiles.html and in the Model Appendix. We model 18 disease states, consisting of the combinations of three stages (T1, T2, T3), three grades (corresponding to Gleason Score 2-6, 7 and > 7), and presence/absence of metastasis. In each of these disease states there are four possible events: progression to a higher disease state, clinical or screen-detection and death. The transition probabilities and durations of different disease states are calibrated to the ERSPC study [16] (model version for Europe) and/or SEER data (model version for US) [6].

After detection, an individual is assigned to either radiation therapy (RT) or radical prostatectomy (RP). In absence of treatment, a baseline survival is assigned at clinical detection, based on data from the pre-PSA era. If an individual is screen-detected, there is a probability of cure that decreases exponentially with lead-time and is calibrated to the observed mortality reduction due to screening in the ERSPC trial [15, 16] (Model Appendix). Each run of MISCAN produces a range of outcomes including, among others, prostate cancer incidence and mortality, life years gained and overdiagnosis [6, 15].

Probabilistic Sensitivity Analyses (ProbSA) using Gaussian Process Regression

The impact of parameter uncertainty on model outcomes is examined by running a ProbSA. A ProbSA consists of repeatedly drawing parameter values from a relevant sampling distribution, and using those to generate an empirical distribution for the outcome of interest. Conducting a ProbSA for microsimulation models of cancer, like MISCAN, is often not feasible, since, we may need many model evaluations to build a reliable empirical distribution of the outcome(s). Therefore we propose to use a Gaussian Process regression model to emulate MISCAN, and minimize the number of MISCAN runs needed for ProbSA.

Gaussian Process Regression

We model the outcome Y, as a function of the input parameters X (defined as an n by p matrix, containing in each row, p by 1 vectors, $x_1, x_2, \ldots, x_i, \ldots, x_n$) as a Gaussian Process (GP). Formally, a GP is a sequence of random variables $Y_1, \ldots, Y_n$, jointly normally distributed,

$$Y_1, \ldots, Y_n \sim N\left(m(X\beta), \sigma^2 C\left(x_i, x_j; \psi\right)\right).$$

We define the mean function $m(.)$ as a simple linear function $X\beta$, and for the covariance function $C(.)$, we choose the commonly used squared exponential [17], which is defined as,

$$C\left(x_i, x_j; \sigma\right) = \frac{\left(x_i - x_j\right)^2}{\psi^2},$$

where ψ is a p by 1 vector of correlation length parameters, which regulates the amount of variation in the outcome due to changes in each of the input parameters.

Step 1: Building a sampling distribution for ProbSA

The first step is to elicit probability distributions for each parameter. We usually calibrate our model using the Nelder-Mead algorithm [18], which does not directly produce confidence intervals.

In general, if there is information about the parameter in the literature we just use its corresponding published confidence interval to determine the level of uncertainty (example: biopsy compliance [19]). If there is no information in the literature, which is the case for most parameters, we derive an empirical confidence interval, based on the distance between observed and predicted data, as measured by a Poisson deviance (Model Appendix). Namely, we vary a parameter (or a block of parameters) until we see an increase in the deviance by more than a certain threshold.

The threshold is based on the 95-th percentile of the chi-square distribution centered at the best-fit value of the deviance. In Table 1 and Supplementary Information Table 1 we show, respectively, the variability associated with each parameter block and the distributions associated with each parameter.

Distributions for each parameter were chosen based on its domain. Parameters which can take any value are assumed to be normally distributed, whereas non-negative parameters (hazard ratios) are assumed to be lognormal distributed, and bounded parameters (probabilities) are assumed to be beta distributed.

Table 1: Summary of included uncertainty in the probabilistic sensitivity analyses*

Parameter Block	Uncertainty Level	Distributions
Onset	0.7%	Beta, Lognormal
Transition Matrix (Odds)	40%	Lognormal
Transition Matrix (Durations)	4%	Lognormal
Extra Clinical Diagnosis US	20%	Lognormal
Hazard Metastasis	10%	Lognormal
PSA growth &	Based on model parameters	Truncnormal
Screening Parameters	Based on Literature	Beta
Baseline Survival	2%	Normal
Effect of Treatment	Based on Literature	Lognormal
Effect of Screening (Cure Parameter)	20% ¥	Beta

* For a complete list with parameter values and their tolerance ranges, see Appendix Table 2. Some parameters were excluded from the overdiagnosis analyses, since they were considered irrelevant a priori, namely all parameters related to survival and some hazard of metastasis parameters. In the column *Uncertainty level*, we show the maximum percentage that a parameter can vary relative to the parameter value. This was determined based on the minimum value that increases the deviance of the incidence fit beyond a threshold based on the 95th-percentile of the chi-square distribution.
& The model for PSA growth is based on an earlier study by de Carvalho *et al* (6) and is calibrated jointly to SEER incidence and ERSPC PSA distribution data.
¥ First the uncertainty in the survival parameters was determined based on the deviance fit of the prostate cancer mortality in the control group of the ERSPC trial. Given these values, the uncertainty in the cure parameter was determined by assessing the deviance between modeled and observed prostate cancer mortality in the screen group of the ERSPC trial (See Appendix Table 1).

Step 2: Choosing model parameters for training

A significant hurdle for the implementation of GP regression for MISCAN lies in the relatively high number of model parameters. For instance, if we would use all 39 parameters (Table 1, Supplementary Information Table 2) to build the emulator, a large number of model runs would be necessary and it would become computationally expensive to obtain an estimate for its parameters, which would limit the emulator's advantage.

Instead, we propose to build a GP emulator on the basis of 10 carefully chosen parameters. We chose the parameters based on two criteria: (1) the parameters are highly influential in the outcome (example: for overdiagnosis, duration in a low-risk disease stage was included, duration in a high-risk disease stage excluded) and (2) if possible, the parameters to include are in different parameter blocks than those already included in the emulator (See Supplementary Information Table 2, for the composition of each parameter block).

Step 3: Training the Gaussian Process Emulator

Given the space spanned by the distributions in Table 1, we build a training dataset, using latin hypercube sampling (R package: `lhs`). We run MISCAN 100 times at the sampled parameter values. The pairs of input parameters and corresponding outcomes are used to estimate parameters β, σ, and ψ of the GP model.
We estimate β and σ using the formulas in [17]. There is no direct formula to estimate ψ. For this we use the same strategy as in [12, 17], which consists of repeatedly plugging in the maximum likelihood estimate for ψ (conditional on the β and σ estimates) in the emulator (using *R* command `optim`).

Step 4: Validation

For validation, we assume that we can obtain an 95% empirical confidence interval of the outcome by running the MISCAN model 1000 times at different sampled parameter values. We define 95% empirical confidence interval as the interval formed by the 2.5-th and 97.5-th percentile of the sorted values of the outcome of interest, either by running MISCAN or the emulator at different sample points.
The discrepancies between the model and emulator outcomes are quantified by computing the average prediction error (as a proportion of the outcome), the standardized individual prediction errors [17] and by comparing the confidence interval obtained with MISCAN and the emulator. We also tested for systematic differences between the emulator an MISCAN using the Mahalanobis distance test [17].

Sensitivity Analyses

Using an emulator instead of the model to perform a ProbSA requires making decisions about the size of the emulator training sample, and choice of parameters. Therefore we conducted several sensitivity analyses to study under which conditions this procedure produces valid results. We use smaller (50) and larger (150) number of MISCAN runs to train the emulator. We run the same procedure including only 5 parameters, instead of 10 and including a randomly chosen set of parameters instead of carefully chosen parameters. We also study by how much prediction error will decrease if the true model only contained the 10 parameters used to train the emulator. Finally we apply the same procedure to examine parametric uncertainty in prostate cancer mortality.

Results

Validation

In Figure 1, the prediction error of the emulator for overdiagnosis is shown. On average, the prediction error equals 1.7% (as a percentage of overdiagnosis). About 97% of the predictions have a prediction error smaller than 5%, and about 37% have an error smaller than 1%. All the standardized individual prediction errors are within their expected values (i.e., smaller than 2 in absolute value, Supplementary Information Figure 2). Despite these favorable outcomes, the value of the Mahalanobis distance statistic is higher than its expected value, which means, there is some discrepancy between the emulator and MISCAN. (Supplementary Information Table 4).

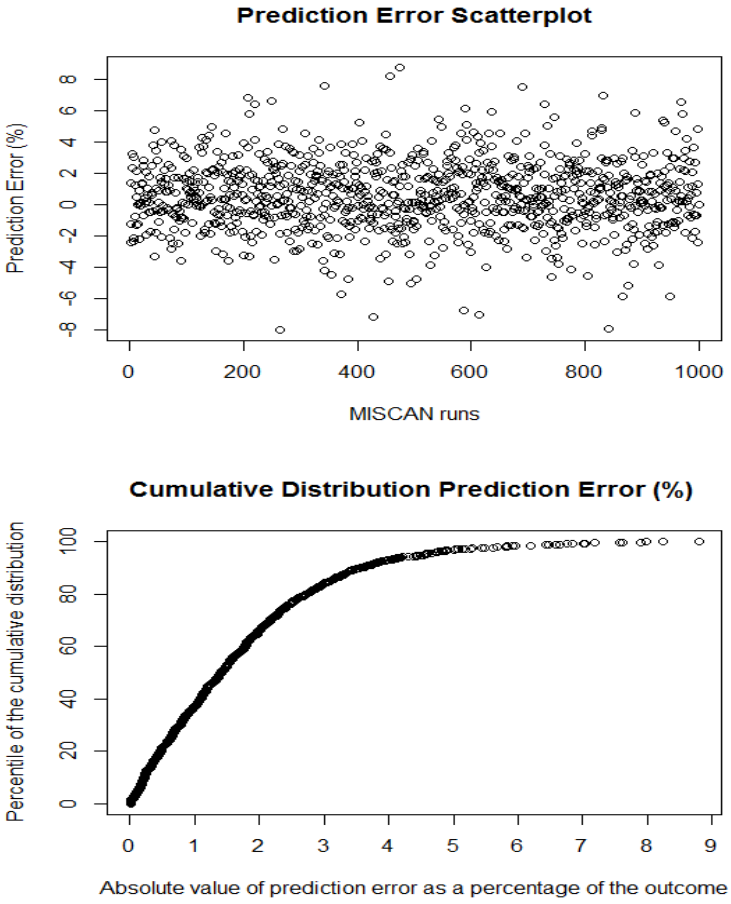
Potential Running Time Savings

The value of using an emulator is dependent on the running time of the particular simulation model. The cost of running a ProbSA with an emulator equals the time needed to produce the training data with MISCAN, plus the time needed to fit the GP model. In Table 2, MISCAN and emulator fitting running times are shown. For a typical MISCAN-prostate model it would take several days to perform a ProbSA, since a single run takes almost 30 minutes. By contrast, fitting a GP emulator for overdiagnosis takes about the same time as a single MISCAN run. Adding more parameters, for the prostate cancer mortality emulator increases the running time to about two standard MISCAN runs. Therefore instead of running MISCAN 1000 times, we would run MISCAN 100 times, plus about the computing time equivalent of two runs. If we carry out an additional 30 runs for validation as in 14, 17 this procedure will result in a reduction of more than 85% in computation time.

Predicting Overdiagnosis

In Figure 2 we show the predicted overdiagnosis with the emulator. For a screen policy of annual screening of men aged 55 to 69, with a PSA threshold for biopsy referral of 4, MISCAN predicts about 42% of screen-detected men are overdiagnosed. Using the GP emulator based on 100 MISCAN training runs, we find that the 95% empirical predicted confidence interval (obtained with 1000 emulator samples) equals (38.0% - 48.0%), which is close to the 95% empirical confidence interval (37.4% - 48.1%), obtained by running MISCAN 1000 times.

Figure 1: Validation of the Gaussian Process Emulator for overdiagnosis.



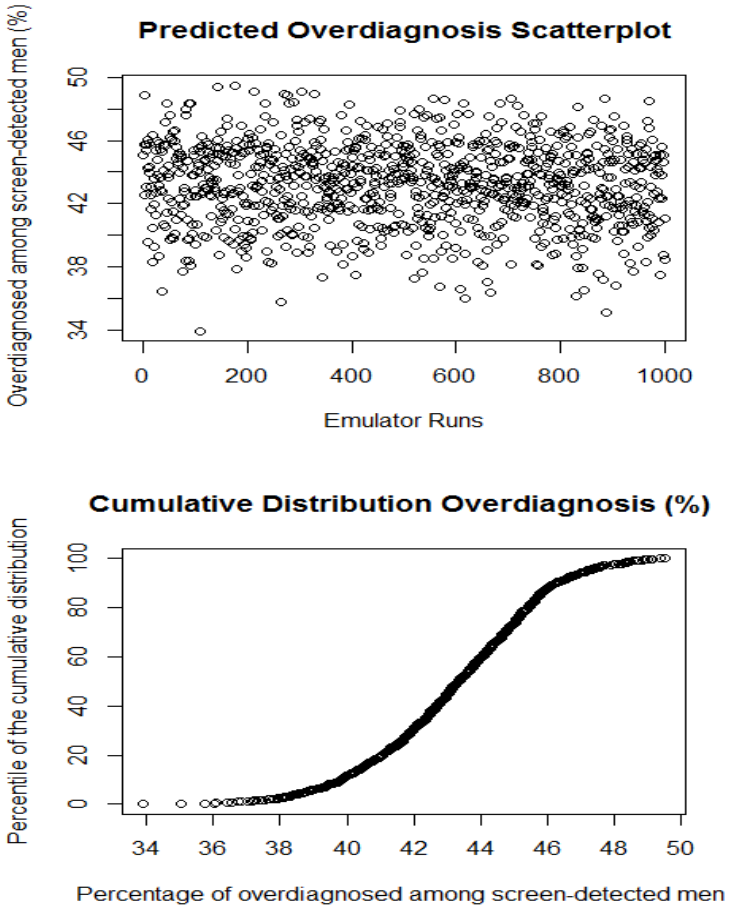
*The panel above contains a scatter plot of the prediction errors. The panel below the cumulative distribution of the absolute predicted errors as a percentage.

Table 2: Running Times of model runs and Gaussian process regression fit.*

	Duration
MISCAN (1 million life histories)	2 min 49 sec.
MISCAN (10 million life histories)	27 min 10 sec.
GP fit and prediction (10 parameters)	18 min 42 sec.
GP fit and prediction (15 parameters)	47 min 28 sec.

* MISCAN is programmed in Delphi (Embarcadero Technologies, Inc.), and all runs were performed in an Optiflex 7010 (Dell Inc.) machine. The Gaussian process emulator was programmed in *R* (R Foundation for Statistical Computing). In this study, the pre-defined sample size was 1 million, since we only use one cohort and in order to make validation feasible. In a typical run we sample 10 million life histories (6,15).

Figure 2: Predicted Overdiagnosis based on Gaussian process emulator.



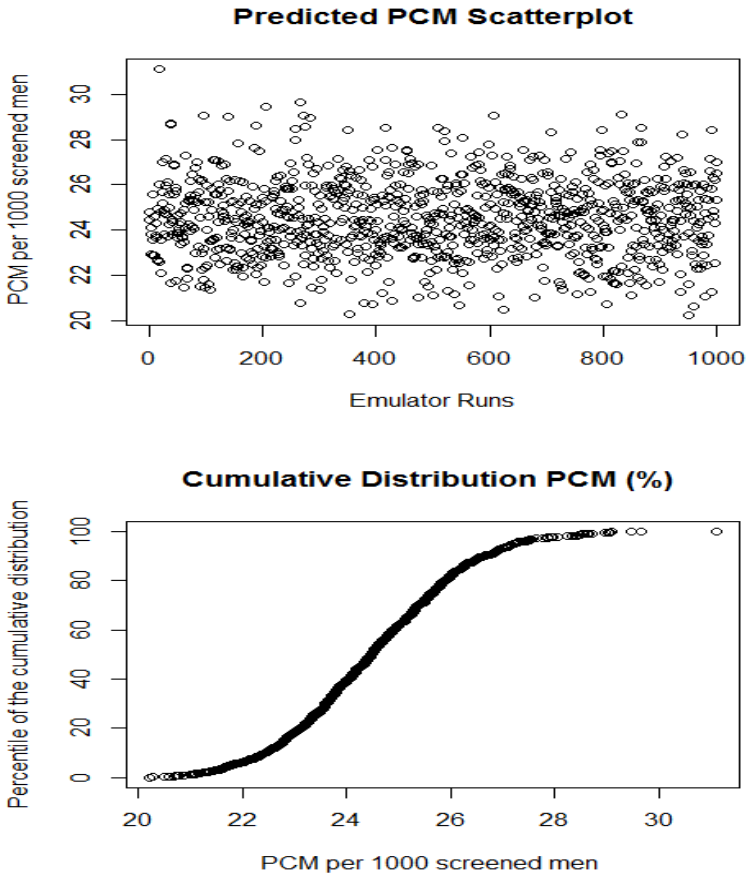
*The panel above contains a scatterplot of predicted overdiagnosis, the panel below the cumulative distribution of overdiagnosis.

Predicting Prostate Cancer Mortality

In Figure 3 we show the predicted prostate cancer mortality with the emulator. In order to build the emulator for prostate cancer mortality we use five extra parameters (i.e. 15 parameters in total). The "true model" contains 12 additional parameters (Supplementary Information Table 2 and 3). For validation, we run MISCAN 50 times and we verify that the average predicted error is about 3%, which is higher than what we found for overdiagnosis. The majority of the standardized individual prediction errors are within their expected values (Appendix Figure 3). The predicted value for prostate cancer mortality per 1000 screened men is 25, and the empirical confidence interval obtained with 1000 samples from the

emulator is (21.4-27.9), which is comparable to the interval found with 50 MISCAN runs (21.6-28.6).

Figure 3: Predicted Prostate Cancer Mortality (PCM) based on Gaussian process emulator.



* The panel above contains a scatter plot of prostate cancer mortality per 1000 screened men. The panel below the distribution of prostate cancer mortality per 1000 screened men.

Sensitivity Analyses

In the sensitivity analyses, we study under which conditions using GP regression will result in a low prediction error. For prostate cancer mortality, we verify that the prediction error increases when fewer MISCAN parameters are included in the emulator. The prediction error did not change significantly by adding or reducing the number of runs to fit the emulator, which means that the 100 runs used to build

the emulator could be excessive. The way that one chooses the parameters to be included in the emulator is important. If we choose them randomly, the average prediction error jumps to about 5%, and the predicted empirical confidence interval with the emulator becomes substantially different from the observed, namely it becomes too small. The same holds, if we would exclude some the parameters that we considered important. Finally, we verify that the average prediction error would decrease to just 0.2%, if in the reference ProbSA (1000 runs with the MISCAN model) we only varied the same 10 parameters as in the emulator.

Discussion

ProbSA are essential to improve the transparency of simulation models and are required by organizations like NICE in the UK [20]. While between-model variability has been analyzed before in numerous Cancer Intervention and Surveillance Modelling Network (CISNET) studies [1-4], few microsimulation studies of cancer screening evaluation assess the impact of parameter uncertainty including all model parameters, like in a ProbSA. This is mostly because it is a computationally expensive procedure and for models like MISCAN, also due to difficulties related to obtaining confidence intervals for the parameters. Previous studies [13, 14, 17, 21] using GP regression to emulate microsimulation models focused on simple models or in a small subset of the model, with at most six input parameters. Our model has more than 50 parameters in total. We have shown that GP regression also works in this context, which is important, since typically, cancer simulation models contain at least ten parameters [1-4].

The computational gain of using a GP regression emulator is dependent on the running time of the simulation model, the number of runs used to train and validate the emulator and the number of parameters included in the emulator. Assuming it is necessary to run the model at least 1000 times to perform a ProbSA, computation time may be reduced by more than 85%, by doing a ProbSA with the help of an emulator. In the sensitivity analyses, we showed that this running time could even be reduced further, by optimizing the number of MISCAN runs to use as data to build the emulator. The prediction error could also be reduced, by optimizing which parameters are included and/or the number of parameters included in the emulator. On the other hand, both of these steps would require additional analyses. The computation time may increase depending on how many validation runs are done with the simulation model.

The performance of the emulator is critically dependent on whether all the important parameters are included. Excluding parameters that may affect every man, instead of a subgroup, or that are expected to affect the disease stages that are

relevant for the outcome of interest, has a significant impact on the emulator's performance. For instance, if the outcome of interest is overdiagnosis, parameters related to low-risk health states are more important than parameters related to the evolution on the disease while in a high-risk disease state, since men in these health states are unlikely to become overdiagnosed. Despite a favorable prediction error, there seem to be some discrepancies in the overall fit, as indicated by a relatively high value of the Mahalanobis distance between the emulator and MISCAN (Supplementary Information Table 4). This is, in principle, due to the fact that we exclude many model parameters, while training the emulator. It could also be due to violations of the assumptions behind the GP model, namely non-normality of the outcome or heteroscedasticity.

MISCAN is calibrated using Nelder-Mead which does not directly produce confidence intervals for each parameter. Our method to determine the uncertainty level for each parameter is based on the difference between the observed data and predicted model output by MISCAN. However, this is just an approximation as we use blocks of correlated parameters, and condition on the values of the other parameters. That is, we implicitly assume that parameters not included in the parameter block (Supplementary Information Table 2) are independent of the included ones, which is likely to be too strong. Consequently, letting parameters vary independently when they are correlated will result in an overestimation of the uncertainty in the outcome. By contrast, models which are calibrated following Bayesian principles using Markov Chain Monte Carlo-like techniques could use the estimated confidence intervals directly in a ProbSA [22-23].

Using GP regression will be most helpful when the simulation model is relatively slow, and with a relatively large number of parameters. For instance, we do not expect that the simulation model in Gulati *et al* [24], which only contains 12 parameters, would need GP regression to evaluate uncertainty, unless its running time would be in orders of magnitude larger than ours.

In conclusion, using a GP regression emulator instead of the model we may reduce the computational effort necessary to carry out a ProbSA by more than 85%, at a cost of a small error. This turns a full ProbSA of a simulation model with a large number of parameters and with a relatively long running time into a feasible task.

References

1. Lansdorp-Vogelaar I, Kuntz KM, Knudsen AB, et al. Stool DNA testing to screen for colorectal cancer in the Medicare population: a cost-effectiveness analysis. Ann Intern Med. 2010;153(6):368-77.

2. Mandelblatt J, van Ravesteyn N, Schechter C, , et al. Which strategies reduce breast cancer mortality most? Collaborative modeling of optimal screening, treatment, and obesity prevention. Cancer 2013; 119(14):2541-8.

3. Meza R, ten Haaf K, Kong CY, , et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. Cancer 2014;120(11):1713-24.

4. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012;118(23):5955-63.

5. Heijnsdijk EA, Wever EM, Auvinen A, , et al. Quality of Life Effects of Prostate Specific Antigen Screening. N Engl J Med 2012;367(7):595-605.

6. de Carvalho TM, Heijnsdijk EAM and de Koning HJ. Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136(7):1600-7.

7. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. Ann Intern Med 2014;160(5):311-20.

8. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. J Urol 2013;190(2):419-26.

9. Wever EM, Draisma G, Heijnsdijk EA et al. How does early detection by screening affect disease progression?: Modelling estimated benefits in prostate cancer screening. Med Decis Making 2011;31(4):550-558.

10. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. J Natl Cancer Inst 2009; 101(6):374-383.

11. Etzioni R, Gulati R, Mallinger L, Mandelblatt J. Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening. Ann Intern Med 2013;158(11):831-8.

12. Kennedy MC, O'Hagan A. Bayesian calibration of computer models. J R Stat Soc B 2001;63(3):425-64.

13. Stevenson MD, Oakley J, Chilcott JB. Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis. Med Decis Making 2004;24(1):89-100.

14. Chang ET, Strong M, Clayton RH. Bayesian Sensitivity Analysis of a Cardiac Cell Model Using a Gaussian Process Emulator. PLoS One 2015;10(6):e0130252

15. 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 2016 . doi: 10.1111/bju.13542.

16. 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(9959):2027-35.

17. Bastos LS, O'Hagan A. Diagnostics for Gaussian Process Emulators. Technometrics 2009;51(4),425-438, DOI: 10.1198/TECH.2009.08019.

18. Nelder JA and Mead R. A simplex method for function minimization. The Computer Journal 1965;7(4):308–313. doi: 10.1093/comjnl/7.4.308.

19. Pinsky PF, Andriole GL, Kramer BS, et al. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol 2005;173(3):746-50.

20. Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14(4):339-47.

21. Becker W Rowson J, Oakley JE, et al. Bayesian sensitivity analysis of a model of the aortic valve. J Biomech 2011 May 17;44(8):1499-506.

22. Boshuizen HC, van Baal PH. Probabilistic sensitivity analysis: be a Bayesian. Value Health. 2009;12(8):1210-4.

23. Rutter CM, Miglioretti DL, Savarino JE. Bayesian Calibration of Microsimulation Models. J Am Stat Assoc 2009;104(488):1338-1350.

24. Gulati R, Inoue L, Katcher J, Hazelton W, Etzioni R. Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. Biostatistics 2010;11(4):707-19.

25. Bill-Axelson A, Holmberg L, Garmo H, ,et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370(10):932-42.

Supplementary Information

Appendix Table 1a: Complete list of parameters and their distributions for probabilistic sensitivity analyses (continues next page)*

Parameter	Value	Distribution	Parameters
Onset **			
Probability of onset	0.51	Beta	(14160.5, 13605.18)
Hazard onset age 30 and	-14.34	LogNormal	(2.66, 0.00)
Hazard onset age 30-50	-6.73	LogNormal	(1.91, 0.00)
Hazard onset age 50-70	-3.46	LogNormal	(1.24, 0.00)
Hazard onset age 70 and older	-1.62	LogNormal	(0.48, 0.00)
Transition Matrix (odds of a transition compared to Clinical Diagnosis)			
Odd T1G6 -> T2G6	3.48	LogNormal	(1.23, 0.20)
Odd T1G6 -> T1G7	2.49	LogNormal	(0.89, 0.20)
Odd T1G7 -> T1G8	2.28	LogNormal	(0.80, 0.20)
Odd T1G7 -> T2G7	3.63	LogNormal	(1.27, 0.20)
Odd T1G8 -> T2G8	-7.67	LogNormal	(2.02, 0.20)
Odd T2G6 -> T3G6	1.72	LogNormal	(0.52, 0.20)
Odd T2G6 -> T2G7	0.99	LogNormal	(-0.03, 0.20)
Odd T2G7 -> T3G7	-12.24	LogNormal	(2.49, 0.20)
Odd T2G7 -> T2G8	0.60	LogNormal	(-0.54, 0.20)
Odd T2G8 -> T3G8	-5.25	LogNormal	(1.64, 0.20)
Odd T3G6 -> T3G7	19.92	LogNormal	(2.97, 0.20)
Odd T3G7 -> T3G8	2.47	LogNormal	(0.88, 0.22)
Transition Matrix (durations in each disease stage)			
T1G6	2.67	LogNormal	(0.98, 0.02)
T1G8	13.68	LogNormal	(2.62, 0.02)
T1G7	4.90	LogNormal	(1.59, 0.02)
T2G6	5.19	LogNormal	(1.65, 0.02)
T2G7	11.12	LogNormal	(2.41, 0.02)
T2G8	20.35	LogNormal	(3.01, 0.02)
T3G6	3.27	LogNormal	(1.18, 0.02)
T3G7	10.58	LogNormal	(2.36, 0.02)
T3G8	20.95	LogNormal	(3.04, 0.02)
Clinical Diagnosis (extra for US population)			
T1 stage	-4.25	LogNormal	(1.44, 0.10)
T2 Stage	-4.15	LogNormal	(1.42, 0.10)
T3 Stage	-3.84	LogNormal	(1.34, 0.11)
Metastasis	5.00	LogNormal	(1.60, 0.10)

Appendix Table 1b: Complete list of parameters and their distributions for probabilistic sensitivity analyses (continued)*

Hazard Metastasis #			
T1G6	3.26	LogNormal	(1.18, 0.05)
T1G7	2.63	LogNormal	(0.96, 0.05)
T2G6	1.72	LogNormal	(0.54, 0.05)
T2G7	2.04	LogNormal	(0.71, 0.05)
PSA growth			
PSA growth without onset	0.02	TruncNormal	(0.33, 0.024)
PSA growth after onset	0.33	TruncNormal	(0.02, 0.001)
Screening Parameters #			
Biopsy compliance	0.41	Beta	(991.38, 1426.62)
Attendance	0.9	Beta	(21.6, 2.4)
Sensitivity T1G6	0.85	Beta	(18.42, 3.25)
Treatment and Screening Benefit ***			
Hazard Ratio RT	0.63	LogNormal	(-0.48, 0.17)
Hazard Ratio RP	0.56	LogNormal	(-0.60, 0.19)
Cure parameter	-0.22	Beta	(60.18, 213.38)
Baseline Survival &			
HR baseline survival correction	0.82	LogNormal	(-0.20, 0.03)
Surv.ageDx	0.14	Normal	(0.14, 0.006)
Surv.ageDxG8	-0.16	Normal	(-0.16, 0.002)
Surv	-4.31	Normal	(-4.31, 0.004)
Surv.FU	0.07	Normal	(0.07, 0.001)
Surv.FU2	-0.01	Normal	(-0.01, 0.000)
Surv M1	-2.06	Normal	(-2.06, 0.004
SurvM1.FU	-0.06	Normal	(-0.06, 0.001)
SurvM1.FU2	-0.01	Normal	(-0.01, 0.000)

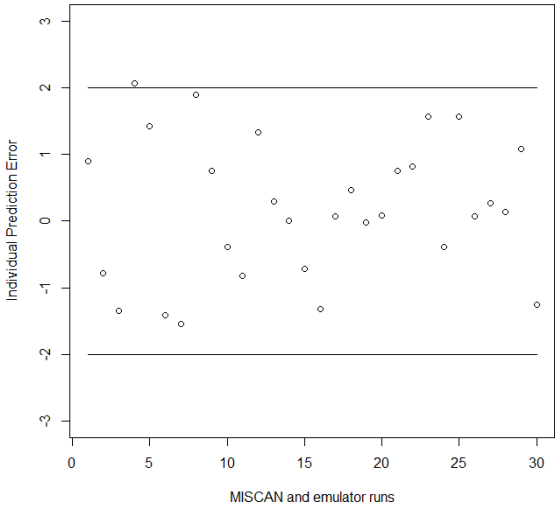
* This corresponds to the models used in (6,15).
** The onset function consists of two components. 1) a probability of onset ; 2) the distribution of onset per age is given by a piecewise constant hazard function dependent on age group.
*** The hazard ratios are relative to the baseline survival. Their value and corresponding distributions are based on the literature (4, 29). The cure parameter is applied on top of the benefit of treatment.
# Some Parameters were excluded.
& HR baseline survival correction (infl.hr) was calibrated to the observed prostate cancer mortality in the control group of the ERSPC trial. The Baseline Survival equation for non-metastatic cases equals,
Prostate Cancer Survival = exp(Surv + Surv.FU*(t-8) + Surv.FU2* (t -8)^2) * infl.hr,
Where t, is the time since onset. By replacing all "Surv" parameters with the set of "Surv.M1" parameters we obtain the survival for metastatic cases. The parameter Surv.ageDx denotes an age specific factor (at the time of clinical diagnosis) applied to the baseline survival.

Appendix Table 2: Parameters included in the Gaussian Process emulator.*#

Parameter Name	Coefficient *(β)*	Correlation Length *(ψ)*
Probability of onset	0.51	9.45
Hazard onset age 30-50	-6.73	6.16
Hazard onset age 50-70	-3.46	20
Duration T1G6	2.67	2.51
Duration T2G6	5.19	4.73
Clinical Diagnosis T1-Stage	-4.2	2.31
Clinical Diagnosis T2-Stage	-4.15	1.4
Clinical Diagnosis T3-Stage	-3.84	1.74
Biopsy Compliance	0.41	1.48
PSA growth after onset	0.33	0.86

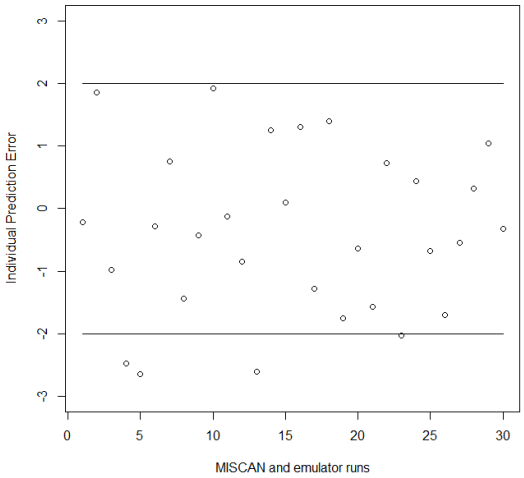
* Contains all parameters included in the emulator. Correlation Length (Phi) parameters were estimated with maximum likelihood estimation. The estimate for the σ^2 parameter equals 4.55 e-05.
# In the 5 parameter run the following parameters were included in the emulator: Probability of onset, Hazard onset age 30-50, Hazard onset age 50-70, Duration T1G6 and PSA growth after onset.

Figure 1: Standardized Individual Prediction Errors for the overdiagnosis emulator.*



*See Figure 2.

Figure 2: Standardized Individual Prediction Errors for the prostate cancer mortality emulator.



* This figure uses data from the first 30 runs (See Appendix Table 4). Individual Prediction Error is based on (18). It equals the difference between the emulator prediction and MISCAN output divided by the standard error of the emulator. A value of the individual prediction error higher than 2 indicates a discrepancy between the emulator and the data, since the standardized prediction errors should be normally distributed.

Appendix Table 3: Mahalanobis distance values for different sensitivity analyses*

Scenario	**Mahalanobis Distance**
Basecase	9900
Another random number seed	11736
N=50	6509
N=150	11373
5 parameters in emulator	> 1 million
Randomly choosen parameters	2773
True model contains 10 parameters	60
Prostate Cancer Mortality Emulator	1269

* Based on the test given in Bastos and O'Hagan (18). The test was calculated using data from the first 30 runs. Expected value of the Mahalanobis distance statistic is 30. Basecase denotes 100 runs, using 123 as a random number seed and 10 carefully chosen parameters. In Another random number seed, seed number 124 is chosen.

Chapter Eleven

General Discussion

General Discussion

This Chapter begins with a description of the answers to research questions followed by the interpretation of the results, limitations and future directions.

Answers to Research Questions

Aim I: Screening

Research Question 1: "Can we find a set of screening policies that significantly reduces the amount of overdiagnosis, while keeping most of the prostate cancer mortality reduction?"

In this study, we estimated overdiagnosis and prostate cancer mortality for about 80 different screening policies, differing by start and stop age, screening frequency, PSA threshold for biopsy referral and policies where screening frequency depends on PSA value.

Using as basecase, yearly screening between ages 50 and 74, we find that we can reduce overdiagnosis by about 30%, just by stopping screening at age 70. This comes at a cost of a prostate cancer mortality increase of about 3%. We can reduce overdiagnosis even further by 60%, if we stop at age 66, but at a cost of about 7% increase in prostate cancer mortality (PCM).

We studied a wide-range of possible ways to perform screening, but we found no evidence of improvement over yearly screening. Restricting screening up to age 70 seems to be a reasonable way to keep the balance of harms and benefits, since increasing the stop age will result in a large amount of overdiagnosis, and the number of lives saved will be limited.

Research Question 2: "Can prostate cancer screening be cost-effective?"

In this study we used ERSPC data and quality of life estimates based on the literature to predict costs and effects of screening for several screening strategies, differing by stop age and screening interval. We found that PSA-screening with intervals shorter or equal than 3 years between PSA-tests are more cost-effective than longer intervals. Using $100,000 as willingness to pay threshold for the incremental cost-effectiveness ratio, we found that screening can be cost-effective if restricted between ages 55 and 59 with a biannual screening interval.

Research Question 3: "What is the influence of comorbidity conditions in the harms and benefits of cancer screening?"

In this study we estimated how the age of screening cessation will vary per comorbidity cohort and cancer site. We used lifetables corresponding to different degrees of comorbidity ranging from, no comorbidity to severe comorbidity, based on Medicare data. We determine the best age to stop screening, by comparing the ratio of number needed to screen per life year gained against a reference ratio of men aged 74 with average health.

For every cancer site, we find that an increased level of comorbidity resulted in a decreased optimal age of cessation of screening. In particular for prostate cancer the cessation age ranged from 78, for people with no comorbidity, down to never screen for men with severe comorbidity.

Research Question 4: "Why is the prostate cancer incidence higher in African-Americans than in the general US population?"

In this chapter we try to ascertain which components of the natural history of prostate cancer explain the higher incidence of prostate cancer for blacks compared to the general US population. We used three independently developed models of prostate cancer natural history using PSA screening patterns by race, based on responses to the National Health Interview Survey in 2005.

We find that African-Americans have a higher chance of developing prostate cancer, and progressing to metastatic disease. On the other hand, given onset of prostate cancer, the chance of detection is similar to the general US population. These findings suggest that PSA-screening should be done in a different way for African-American men than for the general US population, namely with more frequent screening or/and an earlier starting age.

Aim II: Active Surveillance

Research Question 5: " What is the prostate cancer mortality increase and overdiagnosis reduction associated with Active Surveillance, in comparison with Immediate Radical Treatment? "

For this study, we first built a module in MISCAN to simulate Active Surveillance (AS). We estimated probabilities of referral to treatment, given progression of the disease during AS, based on Johns Hopkins cohort data. We projected the lifetime risk of overdiagnosis and prostate cancer mortality of screening combined with different Active Surveillance protocols.

We found that the increase in prostate cancer mortality due to AS (i.e., due to delayed radical treatment) is modest, about 1%. The corresponding overdiagnosis reduction is about 18%. However, by decreasing the frequency of biopsies to 3 years after the first year the overdiagnosis reduction becomes about 30%, with only

a 2% prostate cancer mortality increase. These results show that Active Surveillance is relatively safe and will significantly reduce overtreatment among low-risk men.

Research Question 6: "Do personal characteristics matter when choosing between immediate radical treatment and active surveillance in an older age group (65+)?"

In this study we use comorbidity-specific lifetables, based on MEDICARE data. We estimated the probability of overtreatment and prostate cancer mortality, for men who were screen-detected, by age of screen-detection, disease-stage at detection, comorbidity status and screening history.

We found that disease stage at screen-detection is a critical factor for overtreatment and prostate cancer mortality. While men in the low-risk group have a high probability of being overtreated, and should seriously consider Active Surveillance, men in the high risk group will likely benefit from immediate treatment, unless they have severe comorbidities. For intermediate risk men, age and comorbidity also play an important role, as someone younger and without comorbidity has a risk profile which is similar to high risk men.

Research Question 7: "Is Active Surveillance more cost-effective than Immediate Radical Treatment?"

In this study we used the same model as in Research Questions 5 and 6 to quantify the associated cost and effects of choosing Active Surveillance (AS) instead of immediate radical treatment. We projected the costs and effects of three screening policies, differing by stop age and frequency, combined with three AS protocols with different inclusion criteria and biopsy intervals.

All screening and treatment protocols where immediate treatment was offered for all men were dominated (more costly and resulting in less QALYs gained) or weakly dominated (less QALYs gained but with a higher cost-effectiveness ratio than the next most expensive policy). The same holds true if we refer low and intermediate risk men to AS. Including only low-risk men in AS was cost-effective, when combined with quadrennial screening. It resulted in an incremental cost-effectiveness ratio (ICER) lower than $100,000 per QALY gained. For yearly screening (55-69), the ICER was above $200,000 per QALY gained.

We conclude that AS is more cost-effective than immediate treatment, and that performing AS for low-risk men with triannual biopsies after the first year of follow-up seems to be more efficient. On the other hand, we note that the differences in costs and effects between the different AS protocols are relatively small.

Research Question 8: "When should Active Surveillance for prostate cancer stop if no progression is detected?"

In this study we use the same model as in Research Questions 5, 6 and 7, and the same costs and utility estimates as in Research Question 7, to address a gap in the literature relative to the timing of AS cessation, if no progression is detected. We project the costs and effects of AS protocols differing by the number of follow-up biopsy rounds compared to Conservative Management, where patients can only be treated, if and at the time, when they would be clinically diagnosed in absence of screening.

We conclude that AS is cost-effective compared to Conservative Management for up to 7 biopsy rounds, but only for the age group 55-65. For these men an intensive follow-up biopsy schedule is recommended. For men older than 65, AS is not effective after taking into account quality of life outcomes. For these men there is no need to stay for many years on AS.

Aim III: Parameter Uncertainty in MISCAN

Research Question 9: "Can we evaluate parameter uncertainty in MISCAN in a more efficient manner?"

In this chapter we analysed the effect of parameter uncertainty on overdiagnosis estimates. This is usually performed within the probabilistic sensitivity analyses framework. There are several obstacles which prevent us from using a probabilistic sensitivity analysis as a routine tool with MISCAN. The method used to estimate MISCAN parameters (Nelder-Mead algorithm) does not produce confidence intervals. Additionally MISCAN has more than 50 model parameters, and its running time is somewhat slow. Most parameters in MISCAN are calibrated and they are not observed or there is not much information in the literature about them.

Our solution to evaluate uncertainty is: (1) we use the distance between observed and predicted data, to obtain approximate confidence intervals for model parameters; (2) in order to reduce the model running time, we used an emulator, that is, a statistical model based on data of model inputs and outcomes that mimics the behaviour of MISCAN. This approach resulted in a predicted confidence interval of 37%-48%, for overdiagnosis, and upon validation we verify that the prediction error, due to the usage of the emulator instead of MISCAN, is modest (on average smaller than 2%).

Interpretation of the findings

Towards more effective PSA-based Screening

PSA-based screening saves lives, however it also causes significant harm. The major focus of the screening aim was to study multiple avenues to make PSA-based screening more effective. Current guidelines on how to do screening are contradictory. While the USPSTF issued a recommendation against PSA-based prostate cancer screening other organizations recommend shared-decision making between ages 55 and 69 (1, 2).

The impact of the negative USPSTF recommendation seems to have resulted in a decrease of the incidence among all age groups (3, 4), however, it is difficult to move away from inefficient opportunistic screening, since there is no clear "best" screening protocol. The screening protocol in the ERSPC consisted of a PSA test every 4 years for men aged between 50 and 74 (with a core age group between 55 an 69), while the PLCO screened men aged between 55-74 with PSA every year during 6 years (5, 6). It is unlikely that any of these screening protocols constitute the best balance of harms and benefits of PSA-based screening given the multitude of ways in which screening can be performed.

Modelling is likely the only way through which we can find the best set of protocols, given that it would be impracticable to run a trial with a large enough power to detect an effect of screening between several competing screening protocols. In Research Questions 1-3 we use a microsimulation model validated to ERSPC data (for projecting effects of policies in Europe) and /or SEER incidence data (for the US population), to project the effects of multiple ways of doing screening, differing by age of start and cessation, frequency, PSA threshold and where the frequency of screening depends on PSA value.

In general we find that overdiagnosis increases strongly with age, and in particular strategies where screening age cessation is beyond age 70 are likely to result in a large amount of overdiagnosis. Consequently, this will result in a large reduction in life years gained, when adjusting for quality of life outcomes. Therefore, screening programmes that focus only on younger age groups are more likely to be cost-effective.

For instance, in Chapter 2, overdiagnosis decreased by 30% when we reduced stop age from 74 to age 70, and by 60% when stop age would reduce to age 66. In Chapter 3, we found the average cost per QALY of yearly screening between 55 and 75 was about $320,000. By contrast, screening between 55 and 59, with 2 year interval resulted in an average cost per QALY of $45,615.

The trade-off between prostate cancer mortality and overdiagnosis and the cost per QALY found in Chapters 2 and 3 may not be applicable for African-Americans. As we showed in chapter 5, there is a higher probability of onset and progression of prostate cancer, for African-Americans. Potential implications for these findings

would be to recommend more frequent screening and/or an earlier start age for African-Americans. On the other hand, it is unclear what would be the effect of recommending such policies on the harms and cost-effectiveness.

There are some limitations associated with these results. First, we used a biopsy compliance of 90% in the base case of chapter 1. Biopsy compliance in the US is much lower, about 40% (7), and our sensitivity analyses showed that using such a low biopsy compliance will decrease the efficiency of screening, that is for the same level of overdiagnosis, there will be less lives saved. In chapters 1-4, we use a cure rate based on Wever et al (8). This cure rate is dependent on the disease stage (Gleason ≤ 7 and Gleason > 7) and can only take two values (in Chapter 3, 3 values). By contrast, in chapters 6-9 we use a continuous cure rate which exponentially decreases with lead-time, which is likely to be more realistic (See Model Appendix Table 1 for an overview).

Reducing the harms of screening with Active Surveillance

The main focus of the Active Surveillance (AS) aim is to find a set of screening and AS protocols that will minimize the amount of overtreatment, and the risk of prostate cancer mortality.

Current evidence from clinical cohorts seems to suggest that AS is safe, with a low number of prostate cancer related death events (9). Recent evidence from a randomized control trial comes from the ProtecT trial (10). Their main finding is that there is no significant difference in prostate cancer mortality between treatment and AS arms, after 10 years. However, the way in which AS should be carried out, in order to minimize overtreatment and guarantee safety is still unclear (9, 11-12). Namely, there is no consensus on what the definition of clinically insignificant cancer is, and this results in different selection criteria and/or triggers for intervention while in AS (12). In the US, AS rate of utilization has moved from a mere 10% of low-risk men up to 40% (13) and in some cohorts to 50% (14).

Based on current clinical evidence on the benefit of screening, we only screen men in the 55-69 age group in the model (15). There is considerable uncertainty regarding the best possible AS protocol, therefore, we considered three protocols: including low-risk men with yearly biopsies, low-risk men with a biopsy every three years after the first year and low and intermediate risk-men with yearly biopsies.

Our main results show that the protocol including low-risk men, with triannual biopsies after the first year substantially reduces overdiagnosis (about 30%) with a relatively low increase in prostate cancer mortality (2.3%). It appears this is more efficient, than including intermediate risk men, however, some caution is advised when interpreting results about intermediate risk men in AS, since the probabilities of referral to treatment are based on the Johns Hopkins cohort, which did not include any intermediate risk men. To model this we assume that given progression (which is determined by our natural history model), the probability of referral to treatment is similar to low-risk men.

In chapter 7, we estimated the cost-effectiveness of screening and AS, compared with screening and immediate treatment. In addition to screening yearly between 55 and 69, we studied less intensive screening protocols (given the results in Chapter 3) either by an earlier screening cessation age or with the same stop age but a longer screening interval. Our results show that AS is more cost-effective than immediate treatment after taking into account several combinations of screening and AS protocols. As in chapter 5, it seems that an AS protocol with triannual biopsies after the first year is to be preferred, however, the differences between the different AS protocols are small. Importantly, we found, for a willingness to pay threshold of $100,000 that limited screening (55-69, quadrennial) combined with AS can be cost-effective.

While in Chapters 5-7 we simply assumed that for every man, AS stops at age 75, if no progression is detected, in chapter 8, we studied when AS should be stopped. As far as we know, there is not yet an explicit guideline on this. In practice, this will likely depend on age of screen-detection, and personal characteristics like comorbidity level. We studied this by comparing the lifetime outcomes of a patient, under Conservative Management and Active Surveillance regimens, using the cost-effectiveness analysis framework. Our findings show that, for men younger than 65, AS is cost-effective up to 7 yearly biopsy rounds. For these men, a stricter biopsy schedule is needed, and immediate treatment should be considered. On the other hand, for men older than 65, AS is not effective. Therefore, for these men, there is no need to perform many repeat biopsies.

The main limitation of the AS model is the fact that the probabilities of referral to treatment while in AS given progression, are based solely on the Johns Hopkins cohort treatment-free survival (15). Other cohorts have slightly different inclusion criteria, likely less strict than Johns Hopkins cohort, and different intervals between biopsies (9).

Our model assumes 100% of low-risk men choose AS, while in the US, this number is likely to be under 50% (13, 14). We also did not consider men who progress and choose not to be treated (16). It is likely, that this happens due to considerations which are not fully explicit in the data, like the level of comorbidity.

In chapters 5 and 6, we did not include data on biopsy compliance, since there were no data available in the literature. We assumed this was 90%. Based on PRIAS data (17) we observe that the biopsy compliance is somewhat lower, and decreases with time spent on AS. Our findings in Chapter 8 suggest that our initial assumption on Chapter 5 of stopping AS for every man at age 75 is likely far from optimal. Together, these two limitations will likely result in underestimation of the overtreatment decrease and prostate cancer mortality increase, in Chapter 5.

How to advice a patient given current evidence?

Prostate cancer screening is a controversial subject, given the PLCO and ERSPC debate. Even if we assume that prostate cancer screening saves lives, as shown in ERSPC trial, the probability that a patient may benefit from screening may vary widely from patient to patient.

On the other hand, there is much that we know now: (1) an overwhelming majority of overdiagnosis cases is concentrated in men older than 60, and in men older than 70, screening may be hard to justify (Chapters 2, 3 refs. 18-20). (2) Active Surveillance should be the default option for screen-detected low-risk men (Chapter 5-8, refs 9, 21, 22). (3) Screening could be cost-effective if limited to the age group 55-59 (Chapter 3).

While these findings may give some certainty to doctors and patients, there may be substantial discrepancies on the probability of overtreatment, even within the same age group. For instance, Wever et al (20) found that for men screen-detected between 65 and 69 the probability of overdiagnosis ranged from 9% to 50%, depending on disease stage at detection.

In this thesis, we look at life expectancy as a way to personalize advice on screening. In chapter 4, we find that comorbidity has a strong effect on the age of screening cessation. Namely, assuming that the harms-benefit relation for a reference screening policy for men with average comorbidity at age 74 is the decision-rule for when to screen, men with severe comorbidity should not be screened and men without comorbidities could be screened beyond age 75.

In Chapter 6, we used the AS model together with the comorbidity lifetables, to inform the decision between immediate treatment and AS for men older than 65, which is the group where overdiagnosis has more impact. We find that for low risk men, AS seems to significantly reduce overtreatment at a cost of a modest increase in prostate cancer mortality. For intermediate risk men, the decision between immediate treatment or AS may depend on age and comorbidity status. While some of these men have a profile of a low risk man, especially if older and/or with comorbidities, others have a risk of prostate cancer death closer to the high risk group. Men screen-detected in a high risk disease stage may benefit from immediate treatment even beyond age 70.

A limitation in these studies is that the comorbidity lifetables used are based on MEDICARE data, thus they only start at age 65. Additionally, they are not relative to a specific disease, but to a comorbidity score (23).

The AUA guidelines recommend shared-decision making between age groups 55-69 (15). The burden of screening could be reduced, by screening men with lower PSA's less frequently after age 60, perhaps with only one additional screen after 8 years (24, 25). Men screen-detected in a low-risk stage (Gleason 6, T-stage < T2b, PSA <10) should be referred to AS, until progression occurs (Chapter 6-8), though patients preferences regarding treatment should also be taken into account.

This policy may need to be fine-tuned, for men with family history or African ancestry, which could start screening earlier and more frequently (Chapter 5), and men with significant comorbidities, which could have an earlier age of cessation (Chapter 4, 6).

It is crucial to prevent opportunistic screening from occurring in later ages (>75) (Chapters 2, 3 and 7, refs. 3, 18-20), and if men decide to be tested, to ensure that only those who need treatment (Gleason 7 or higher, stage T2 or higher, no or mild comorbidities) are radically treated.

Uncertainty in Simulation Models

Results of microsimulation models are subject to uncertainty. We usually distinguish between three types of uncertainty. Simulation uncertainty, which is due to random errors, can be minimized by simulating a large enough cohort. Parametric uncertainty is the uncertainty in the model outcomes due to model parameters and structural or model uncertainty is the uncertainty in the outcomes due to the uncertainty in assumptions used to build the model.

In Chapter 9, we analysed the effect of parametric uncertainty on the overdiagnosis estimates. This is usually performed within the probabilistic sensitivity analyses framework. Given the fact that this procedure is computationally expensive, we used an emulator for the model, based on Gaussian Process Regression. With this we showed it becomes feasible to obtain a confidence interval for model outcomes. There is also uncertainty associated to assumptions used to build the model. We routinely collaborate with other prostate cancer screening modelling groups under the CISNET framework, where independently developed models using common data sources are compared in a systematic way (26).

For instance, a study including three models (including MISCAN) found a range for overdiagnosis between 23% and 42% for screen-detected men (27). In chapter 5, we estimated that the lifetime probability of onset for African-Americans is between 31% and 45%, across three independently developed models. In a comparable study to chapter 2, by Gulati et al (28), where prostate cancer mortality and overdiagnosis were projected for different screening policies, it seems that for the same level of overdiagnosis, there is less prostate cancer mortality. Previous MISCAN studies have also looked at the effects of changing several model assumptions (20, 29).

The utilities used for studies where QALYs are the primary outcome, are subject to substantial uncertainty. In our experience, one of the critical parameters is the utility of the post-recovery period, after radical treatment (Model Appendix Table 11). This utility determines how much harm the treatments cause, and we assume a relatively long duration (the utility decrement holds on for 9 years after treatment). One source of uncertainty comes from the quality of life of patients treated in low and high volume medical centres, since men treated in high volume centres are likely to have less complications and side-effects (30).

There is also uncertainty about the discounting rate used. There are several guidelines on which discount rate to use and these may differ substantially across countries. For instance, UK guidelines recommend discount rate of 3.5% for cost and effects (31), while The Netherlands recommends a 1.5% discount rate for effects and 4% for costs (32). Therefore some caution is advised when translating cost-effectiveness results between countries.

Future Directions

Risk Stratification

PSA-screening can be made more effective, with better risk-stratification. Risk stratification based on the value of PSA at age 60 (studied in Chapter 1) seems to be a promising avenue to decrease the burden of screening (24, 26). This way, about 50% of men could have a less intensive screening schedule or even stop after age 60. Still, this would not solve the "conundrum", as few men with PSA lower than 1, would have a PSA higher than 3 at a later point, that is, not many biopsies would be avoided (See Model Appendix Table 4).

A limitation of the PSA test is that it is not prostate cancer specific (33). Elevated PSA could be caused by other benign conditions. This results in many unnecessary biopsies that need to be performed, decreasing the quality of life of the men who undergo PSA-screening. Therefore, many efforts have been devoted to find better biomarkers (34, 35), genetic markers (35, 36), risk calculators (37), or a better combination of biomarkers clinical variables and genetic markers (38). Other significant efforts are being made with novel imaging technologies like MRI (39, 40). Each of these markers or technologies could be applied at different stages of screening, namely, before or together with PSA test, after the PSA and before biopsy and finally to select patients with lower likelihood of developing advanced disease for Active Surveillance.

Examples of two biomarkers are PHI and PCA3. PHI is a combination of three PSA derivatives. It was shown that PHI has a higher AUC and could reduce the number of biopsies by about 16% (33). PCA3 (prostate antigen 3) is a marker which is overexpressed in prostate cancer tissue, which also has shown some predictive value (34). As of 2013, more than 70 prostate cancer susceptibility loci have been identified. These explain about 30% of family history risk, and for the top 1% the risk distribution, there is a 4.7-times higher risk of prostate cancer than for the average men (35, 36). This type of information could be used to develop a personalized screening schedule. There seems to be more value in combining several biomarkers and other clinical variables of interest, than by using a single one. In Poyet et al (37) the usage of risk calculators increased AUC from 0.58 (with PSA only) to 0.65-0.66 and performed even better for clinically significant prostate cancer. The Stockholm-3 trial (38) studied a test which combined PSA, other biomarkers, genetic markers and clinical variables, and found that the AUC improved from 0.56 to 0.74. This resulted in about 32% less biopsies needed. MRI-targeted biopsy seems to be more sensitive for high risk cancers, with substantially

more high risk cancers detected compared to standard biopsy, while at the same it diagnosis less low-risk cancers (39, 40).

Questions about how much overdiagnosis or prostate biopsies will be reduced, and the cost-effectiveness of MRI, biomarkers or combinations of markers, could potentially be answered by MISCAN in the future by modulating the overall test and biopsy sensitivity and specificity.

However, there are some methodological issues relative to the performance of biomarkers, due to uncertainty in the sampling of the prostate tissue at biopsy. If the biopsy found no cancer, it could be a false negative as its sensitivity is not perfect (39), and therefore evaluation of biomarker performance is difficult. Regarding MRI, its performance may also depend on the operator (39). It is also uncertain how to best combine all these variables, and which threshold scores for biopsy referral should be used.

Following the results of Chapter 4, on the racial differences between black and white men on the prostate cancer natural history, it seems that management of prostate cancer care should be done in a different way for the African-American population. MISCAN could be used to project the harms and benefits of different screening and AS protocols for black men. The main obstacle for this is to obtain high quality data on prostate cancer survival for black men. For instance, the benefit of treatment in MISCAN is based on a Swedish cohort (41), and the benefit of screening is based on ERSPC trial (6), which are both unlikely to contain a significant African population.

Active Surveillance

While there is an emerging consensus on the referral of low-risk men to AS, there is substantial uncertainty regarding its application (9-12). A significant proportion of men in AS are eventually referred to treatment. There are several reasons for this, for example, the fact that biopsy of the prostate is not sensitive enough, namely many Gleason 6 cancers could actually be Gleason 7 cancers which were misidentified. Some patients also leave AS, without a protocol based reason, due to anxiety (17). AS could become more cost-effective, by employing stricter selection with the help of MRI and/or biomarkers and risk calculators. If clinical cohorts keep showing low rates of prostate cancer related death events, it is likely that the uptake of AS will continue to increase.

General Conclusions

In this thesis, my aim was to quantify the harms and benefits of prostate cancer screening and treatment. My findings can be summarized as follows,

(1) **PSA-based prostate cancer screening** can be made more efficient by focusing on younger age groups. Almost all overdiagnosed cases of prostate cancer occur in men older than 60. After age 70, though a few lives are saved, the amount of overdiagnosis added with each screening round becomes overwhelming. Screening may be cost-effective if restricted to men younger than 60.

(2) **Active Surveillance for prostate cancer** should be the default option for low-risk men older than 60. Additionally, we find that doing triannual biopsies seems better at reducing overtreatment and more cost-effective, than doing annual biopsies, though the differences between AS protocols are small. For intermediate risk men, the choice between immediate treatment and AS, may depend on personal characteristics like age and level of comorbidity. Men older than 65 don't need to stay many years on AS, if no progression is detected.

(3) **Personalized screening strategies** are needed to improve the current balance of harms and benefits of screening. For instance, we found that Black men have a higher risk of onset than Caucasians, which may indicate a need to start screening earlier. Men with significant comorbidities have a significantly higher probability of not benefitting from screening and treatment, which may indicate an earlier stopping age for screening than for the general population.

(4) **Uncertainty in the results of simulation models** can be analysed with probabilistic sensitivity analyses, even for large models like MISCAN, by using statistical techniques to emulate the model, like Gaussian Process Regression.

Recommendations

(a) Start PSA-screening pilot studies, including men aged between 55 and 60.

(b) Avoid PSA testing in men older than 70.

(c) Present Active Surveillance low-risk men older than 60 as the default treatment approach.

(d) Increase the role of risk stratification at every stage of the screening process.

(e) Build a web-based tool, including MISCAN model estimates, for clinicians and patients.

(f) Publish future model estimates including an uncertainty interval. If this is a computationally expensive task, then use an emulator (like Gaussian process regression).

References

1. Moyer VA; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012 17;157(2):120-34.

2. Loeb S. Guideline of guidelines: prostate cancer screening. BJU Int. 2014 ;114(3):323-5.

3. Jemal A, Fedewa SA, Ma J, et al. Prostate Cancer Incidence and PSA Testing Patterns in Relation to USPSTF Screening Recommendations. JAMA. 2015;314(19):2054-61.

4. Drazer MW, Huo D, Eggener SE. National Prostate Cancer Screening Rates After the 2012 US Preventive Services Task Force Recommendation Discouraging Prostate-Specific Antigen-Based Screening. J Clin Oncol. 2015 Aug 1;33(22):2416-2.

5. Andriole GL, Crawford ED, Grubb RL, et al. Mortality Results after 13 years of Follow-up. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med 2009; 360:1310-19.

6. 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.

7. Pinsky PF, Andriole GL, Kramer BS, et al. Prostate, Lung, Colorectal and Ovarian Project Team. Prostate biopsy following a positive screen in the prostate, lung, colorectal and ovarian cancer screening trial. J Urol. 2005;173:746-50.

8. Wever EM, Draisma G, Heijnsdijk EA et al. How does early detection by screening affect disease progression?: Modelling estimated benefits in prostate cancer screening. Med Decis Making. 2011;31(4):550-558.

9. Hamdy FC, Donovan JL, Lane JA, et al. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med. 2016;375(15):1415-1424.

10. Simpkin AJ, Tilling K, Martin RM, et al. Systematic Review and Meta-analysis of Factors Determining Change to Radical Treatment in Active Surveillance for Localized Prostate Cancer. Eur Urol 2015;67:993-1005.

11. Klotz L. Active Surveillance for Prostate Cancer: Debate over the Application, Not the Concept. Eur Urol. 2015;67(6):1006-8.

12. Bruinsma SM, Bangma CH, Carroll PR, et al. Active surveillance for prostate cancer: a narrative review of clinical guidelines. Nat Rev Urol. 2016 Jan 27. doi: 10.1038/nrurol.2015.313

13. Cooperberg MR, Carroll PR. Trends in Management for Patients With Localized Prostate Cancer, 1990-2013. JAMA 2015;314:80-2.

14. Womble PR, Montie JE, Ye Z, et al. Michigan Urological Surgery Improvement Collaborative. Contemporary Use of Initial Active Surveillance Among Men in Michigan with Low-risk Prostate Cancer. Eur Urol 2015; 67:44-50.

15. Carter HB, Albertsen PC, Barry MJ, et al. Early detection of prostate cancer: AUA Guideline. J Urol. 2013;190(2):419-26.

16. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

17. Bokhorst LP, Alberts AR, Rannikko A, et al. PRIAS study group. Compliance Rates with the Prostate Cancer Research International Active Surveillance (PRIAS) Protocol and Disease Reclassification in Noncompliers. Eur Urol. 2015;68(5):814-21.

18. Gulati R, Inoue LY, Gore JL, et al. Individualized estimates of overdiagnosis in screen-detected prostate cancer. J Natl Cancer Inst 2014;106:djt367.

19. Vickers AJ, Sjoberg DD, Ulmert D, et al. Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen. BMC Med 2014;12:26. doi: 10.1186/1741-7015-12-26.

20. Wever EM, Hugosson J, Heijnsdijk EA, et al. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. Br J Cancer 2012;107:778-84.

21. Bangma CH, Valdagni R, Carroll PR, et al. Active surveillance for low-risk prostate cancer: developments to date. Eur Urol. 2015;67(4):646-8.

22. Xia J, Trock BJ, Cooperberg MR, et al. Prostate cancer mortality following active surveillance versus immediate radical prostatectomy. Clin Cancer Res 2012;18:5471-8.

23. Cho H, Klabunde CN, Yabroff KR, et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. Ann Intern Med 2013;159:667-76.

24. Vickers AJ, Lilja H. Predicting prostate cancer many years before diagnosis: how and why?. World J Urol. 2012;30(2):131-135.

25. Roobol MJ, Roobol DW, Schroder FH. Is additional testing necessary in men with prostate-specific antigen levels of 1.0 ng/mL or less in population-based screening setting?. Urology. 2005;65(2):343-346.

26. CISNET. http://surveillance.cancer.gov/publications/factsheets/CISNET_Fact_Sheet.pdf Acessed 17/02/2016.

27. Draisma G, Etzioni R, Tsodikov A, et al. Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. J Natl Cancer Inst 2009; 101:374-83.

28. Gulati R, Gore JL and Etzioni R. Comparative Effectiveness of Alternative Prostate-Specific Antigen–Based Prostate Cancer Screening Strategies. Ann Intern Med. 2013; 158:145-53.

29. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012;118:5955-63.

30. Trinh QD, Bjartell A, Freedland SJ, et al. A systematic review of the volume-outcome relationship for radical prostatectomy. Eur Urol. 2013;64(5):786-98.

31. NICE. Guide to the methods of technology appraisal 2013. https://www.nice.org.uk/article/pmg9/chapter/the-reference-case . Acessed 13/02/2016.

32. RIVM. Op weg naar maatschappelijke kosten-batenanalyses voor preventie en zorg. http://www.rivm.nl/dsresource?objectid=rivmp:246026&type=org&disposition=inline&ns_nc=1 . Acessed 19/02/2016.

33. Lazzeri M, Haese A, de la Taille A, et al. Serum isoform [-2]proPSA derivatives significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml: a multicentric European study.

34. Wei JT, Feng Z, Partin AW, et al. Can urinary PCA3 supplement PSA in the early detection of prostate cancer? J Clin Oncol. 2014;32(36):4066-72.

35. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet. 2013;45(4):385-91, 391e1-2. doi: 10.1038/ng.2560.

36. Eeles R, Goh C, Castro E, et al. The genetic epidemiology of prostate cancer and its clinical implications. Nat Rev Urol. 2014;11(1):18-31.

37. Poyet C, Nieboer D, Bhindi B, et al.Prostate cancer risk prediction using the novel versions of the European Randomised Study for Screening of Prostate Cancer (ERSPC) and Prostate Cancer Prevention Trial (PCPT) risk calculators: independent validation and comparison in a contemporary European cohort. BJU Int. 2015. doi: 10.1111/bju.13314

38. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. Lancet Oncol. 2015;16(16):1667-76.

39. Siddiqui MM, Rais-Bahrami S, Turkbey B, et al. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. JAMA. 2015;313:390-7.

40. Pokorny MR, de Rooij M, Duncan E, et al. Prospective study of diagnostic accuracy comparing prostate cancer detection by transrectal ultrasound-guided biopsy versus magnetic resonance (MR) imaging with subsequent MR-guided biopsy in men without previous prostate biopsies. Eur Urol. 2014;66(1):22-9.

41. Bill-Axelson A, Holmberg L, Garmo H, et al. Radical prostatectomy or watchful waiting in early prostate cancer. N Engl J Med 2014;370:932-42.

Chapter Twelve

Model Appendix

Overview of Model Characteristics *per* Chapter

Table 1: Model characteristics per Chapter.

	Topic	Data	Detection	Survival &	AS	Other
Aim	Screening					
Ch. 2	Screening policies for prostate cancer.	US	Separate PSA growth, biopsy compliance and sensitivity.	Two cure rates for, Gleason ≤ 7 and Gleason > 7	Not modelled.	-
Ch. 3	Cost-Effectiveness of Screening Policies.	Europe	A single joint sensitivity parameter per health state.	Three cure rates for, Gleason < 7, Gleason 7 and and Gleason > 7.	30% of men on AS receive Radical Treatment within 7 years.	-
Ch.4	Screening Cessation based on comorbidity.	US	A single joint sensitivity parameter per health state.	Two cure rates for, Gleason ≤ 7 and Gleason > 7	Not modelled.	Uses comorbidity -specific lifetables.
Ch. 5	Prostate Cancer in black man.	US (by race)	A single joint sensitivity parameter per health state.	Two cure rates for, ≤ Gleason 7 and Gleason 8 or higher	Not modelled.	Race-specific Natural History Parameters
Aim	Active Surveillance					
Ch. 6	Risks and Benefits of AS protocols	US	Separate PSA growth, biopsy compliance and sensitivity	Exponentially decreasing with time to clinical diagnosis.	See Tables 8-9	
Ch. 7	Individual benefit of AS	US	Separate PSA growth, biopsy compliance and sensitivity.	Exponentially decreasing with time to clinical diagnosis.	See Tables 8-9	Uses comorbidity -specific lifetables.
Ch. 8	Cost-Effectiveness AS	US	Separate PSA growth, biopsy compliance and sensitivity.	Exponentially decreasing with time to clinical diagnosis.	See Tables 8-10	Biopsy Compliance based on PRIAS
Ch. 9	When to Stop AS?	US	Separate PSA growth, biopsy compliance and sensitivity.	Exponentially decreasing with time to clinical diagnosis.	See Tables 8-10	Biopsy Compliance based on PRIAS
	Uncertainty in MISCAN					
Ch.10	Evaluating uncertainty with emulators	US	Separate PSA growth, biopsy compliance and sensitivity	Exponentially decreasing with time to clinical diagnosis.	Not Relevant	Uses an R build programme to emulate MISCAN

*AS denotes Active Surveillance.
& Mainly, the differences regarding modelling of survival for screen-detected men.

MISCAN: Model Description

Microsimulation Screening Analysis (MISCAN) is a microsimulation model, which simulates individual life histories including the natural history of prostate cancer. Its main purpose is to simulate prostate cancer screening and treatment policies (for more details see http://cisnet.cancer.gov/prostate/profiles.html).

Natural History

The natural history part is divided into several health states. We model 18 detectable disease stages, consisting of the combination of 3 stages (T1, T2, T3), 3 grades (which correspond to Gleason Score 2-6, 7 and 8-10) and whether or not the cancer is metastasized. In each of these disease stages, an individual can progress to a higher disease state, be clinically or screen detected, become metastasized or can die. Progression is defined by a matrix of odds of moving between states (depending on the particular disease state) and dwelling time distributions for the time spent in each state. The dwelling times are determined by Weibull distributions (1, 2).

Given the probability of onset of prostate cancer in the general population, the age distribution of the onset of prostate cancer is given by a piecewise constant hazard function which increases with age and is defined as,

$$\lambda_o(t) = \exp(\psi_t),$$

where t equals ages 30, 50 and 70.

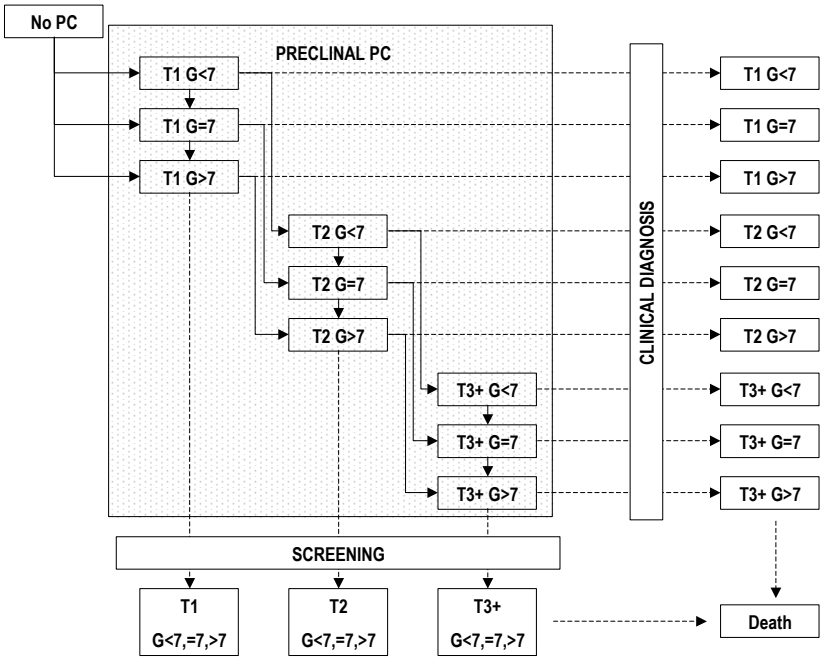
Detection

At each disease state (See Figure 1), a person can be clinical or screen-detected. The durations and hazards of clinical diagnosis depend on the disease state. In absence of screening, metastasis and death other causes, the individual will transition to a clinical diagnosed state if the generated random number is smaller than the probability of clinical diagnosis,

$$p(c_{currstate}) = 1 - p(n_{currstate}),$$

Where *p(.)* denotes probability, *currstate* denotes the current disease state (See table 1) *c* denotes clinical diagnosis event and *n* the event of moving to a higher disease state than *currstate*. The exact $p(c_{currstate})$ is more complex as it depends also on, the hazard of metastasis (if metastasis occurs then there is a higher probability of clinical diagnosis), the screening schedule and other cause survival.

Figure 1: The MISCAN model*



* Reproduced from Heijnsdijk et al (3), Prostate cancer develops from no prostate cancer via one or more screen-detectable preclinical stages to a clinically diagnosed cancer or screen detected cancer. The arrows indicate the possible transitions. Each state can be local or metastatic, but for simplicity this is not illustrated.

The screening part contains several parameters. Conditional on the individual's current prostate cancer disease stage, the probability of screen-detection ($d_{currstate}$) given the current disease state is a function of the probability of attendance (a) , biopsy compliance (b) and the joint PSA and biopsy sensitivity ($s_{currstate}$),

$$p(d_{currstate}) = a^* \; b \; * \; s_{currstate} .$$

The screening schedules in the US are generated based on Mariotto et al (4). The estimation of $s_{currstate}$ parameters is described on Wever et al (5). The parameter $s_{currstate}$ was later changed, and we now model PSA growth explicitly, see section on modelling PSA growth.

Cure Rate and benefit of treatment

The probability of cure is dependent on Gleason Score and the presence or absence of metastasis (this was changed to a lead-time dependent cure see section on Lead-Time Dependent Cure). The estimation process is extensively described in Wever et al (1). In the case where the patient is referred to radical treatment, there is an additional hazard ratio of cure, which depends on the type of treatment. A detailed description of the assumed hazard ratios of cure is described on Etzioni et al (6).

Estimation

Typically, model parameters for the natural history component and the screening tests are estimated as follows: A model is constructed for a specific case, such as prostate cancer incidence in the US or both arms of the ERSPC trial Rotterdam. Parameters are calibrated by minimizing the distance between observed numbers of cases and the cases predicted by the model. This distance is calculated assuming a Poisson deviance function for incidence data. For the minimization an adapted version of the simplex optimization method of Nelder and Mead is used. Optimization is initiated with small sample sizes and repeated with larger sample sizes, until it is not possible to decrease the deviance. (See http://cisnet.cancer.gov/prostate/profiles.html)

The MISCAN-model was primarily validated to the ERSPC trial data. Subsequently it was adjusted for the US situation by adapting the population and the PSA testing practice, namely by inserting US lifetables and screening dissemination patterns based on Mariotto et al (4). Additionally, we included an extra stage-specific hazard of clinical diagnosis, which implies an earlier diagnosis of prostate cancer in the absence of screening in the US population compared to the ERSPC trial. The model is calibrated to the SEER incidence from 1975 to 2000, as well as stage distribution data. The probability of cure was calibrated based on the observed 29% mortality reduction in the ERSPCtrial (Wever et al (1) and Schroder et al (7)).

Modelling PSA growth

We generate a new PSA measurement *j* for an individual *i* as,

$$psa_{ij} = exp\left\{b0 + b1_i\ (age - 40) + b2_i\ (age - onsetage) + agefactor\ (62 - age) + b3_{ij}\right\},$$

Where $b0$ is the constant term , $b1_i$ and $b2_i$ are truncated normal distributed and $b3_{ij}$ is a normally distributed error term.

Table 2: Parameters of the PSA growth equation and biopsy sensitivities

Parameter	*Value*	*Note*
b_0 (Constant Term)	-0.47	Exp(-0.47) matches the average PSA at age 40.*
Age factor	0.07	Centred at age 62. Makes PSA grow slower for older men.
b_{1i} (Age - 40)	0.02**	
b_{2i} (Age – Onset Age)	0.33**	It becomes non-zero, at the time of onset of prostate cancer.
error term (b_{3ij}) variance	0.35	Matches the interquartile PSA distribution at age 40.*
correlation between PSA measurements ***	0.75	Has an effect in individual PSA dynamics. See Table 2.
Biopsy Sensitivity	Value	
T1GS6,T1GS7	0.85	Stage T1 and GS ≤ 7
T2GS6, T2GS7	0.95	Stage T2 and GS ≤ 7
T1GS8, T2GS8	0.97	Stage $\leq$ T2 and GS $\geq$ 8
T3GS6, T3GS7	0.97	Stage T3 and GS ≤ 7
T3GS8	0.99	Stage T3, GS ≥ 8
M1T1	0.97	Metastasis, Stage T1
M1T2	0.98	Metastasis, Stage T2
M1T3	0.99	Metastasis, Stage T3

* Based on Vickers et al. (8)
** This value corresponds to the mean of the distribution.
*** Correlation between the errors (b_{3ij} in the PSA growth generator).

Table 3: Cumulative PSA distribution at 1st Screen (%)

PSA	*Observed (ERSPC)*	*Predicted (MISCAN)*
0.5	14.9	12.1
1	40.2	40.3
1.5	57.4	60.8
2	68.5	73.5
2.5	75.4	81.5
3	80.3	86.8
4	87.0	92.4
5	91.1	95.2
7	95.3	97.7
10	97.8	99.0
100	99.9	100.0

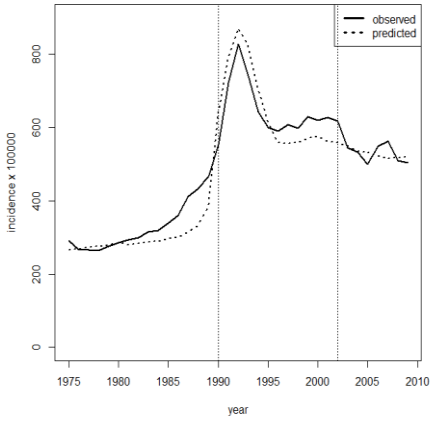
* ERSPC stands for European Randomized Screening for Prostate Cancer Trial. Data corresponding to the first round of screening was used.

Table 4: PSA within individual dynamic

	ERSPC	**Correlation between PSA measurements**				
		0	*0.7*	*0.75*	*0.8*	*1*
Events		*Probabilities*				
PSA < 1 at age 60 and PSA > 3 at age 68	2.2	9.5	3.6	2.3	1.6	1.4
PSA < 1 at age 60 and PSA < 3 at age 68	97.8	90.5	96.4	97.7	98.4	98.6
PSA > 1 at age 60 and PSA > 3 at age 68	35.5	24.1	27.6	28.2	28.8	29.6
PSA > 1 at age 60 and PSA < 3 at age 68	65.5	75.9	72.4	71.8	71.2	70.4

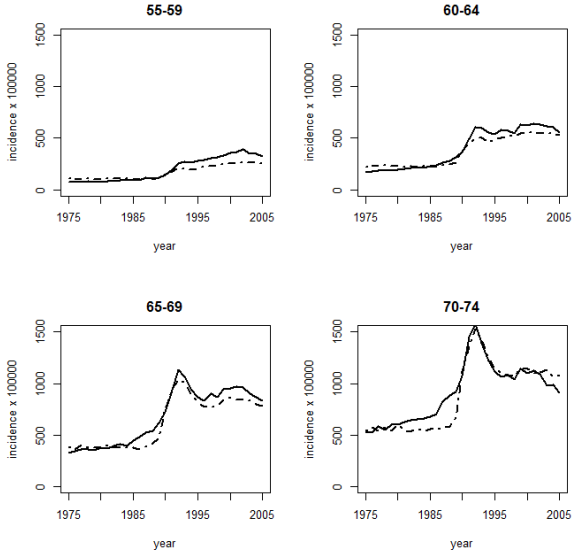
***** This validation is necessary for the screening policies where stop age and frequency depend on PSA value at a certain age. In the ERSPC data (7), men were selected that participated in the first/second screening round about age 60 and that two screening rounds later were still screened around age 68.

Figure 2: Incidence of PC in the US population, between 1975 and 2009, for age group 50-85. (Between the vertical dots is the calibration period, 1990-2002).



*Observed denotes the observed prostate cancer incidence rate in age group 50-84, in the US population based on SEER 1975-2009 data. Predicted denotes the predicted incidence by MISCAN, based on a sample of 10 million men representative of the US age distribution across time. The data used for calibration of PSA growth parameters were 1990-2002 SEER data. Before 1990, the model is similar to Wever et al (1).

Figure 3: Observed and Predicted Incidence in the US population, 1975-2005 per age group.*



* Based on de Carvalho *et al* (9). Observed denotes the observed prostate cancer incidence rate in age group 50-84, in the US population based on SEER 1975-2005 data. Predicted denotes the predicted incidence by MISCAN, based on a sample of 10 million men representative of the US age distribution across time.

Lead-Time Dependent Cure

Table 5: Observed and Predicted cumulative prostate cancer deaths in the ERSPC*

	Observed	MISCAN
ERPSC-Control	415	415
ERSPC-Screening	265	266

*Using data at 11 years of follow-up (7). The survival in absence of treatment (based on pre-PSA era data) was corrected with a hazard ratio calibrated to the observed prostate cancer mortality in the control group of the ERSPC trial, which equals 0.82. The effect of screening is calibrated to the observed prostate cancer mortality reduction after 11 years of follow-up in the ERSPC trial.

Table 6: Observed and Predicted cumulative prostate cancer deaths in the ERSPC (13-year follow-up) by age at randomization.*

	ERSPC-Control		ERSPC-Screening	
	Observed	Predicted	Observed	Predicted
55-59	114	113	174	187
60-64	121	123	159	192
65-69	120	128	212	206
Total	355	364	545	585

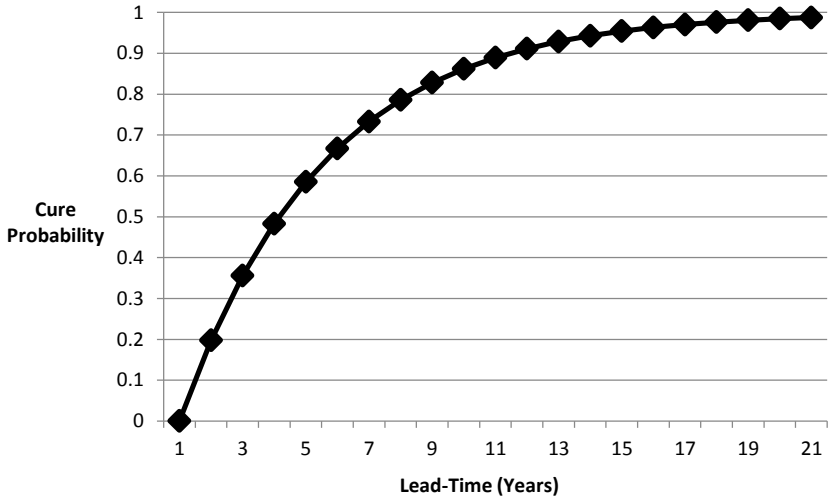
* Using data at 13 years of follow-up (10). In the ERSPC the follow-up is not complete, but it is complete in MISCAN. For this reason we calibrated the cure parameter to the observed prostate cancer mortality reduction after 11 years of follow-up in the ERSPC trial. The survival in absence of treatment (based on pre-PSA era data) was corrected with a hazard ratio calibrated to the observed prostate cancer mortality in the control group of the ERSPC trial, which equals 0.82.

Table 7: Sensitivity Analyses, Observed and Predicted cumulative prostate cancer deaths in the ERSPC &

Trial Arm	Observed	Linear Fit	20% less cure	5% higher HR
Control	415	415	415	439
Screening	265	308	276	276

& "Linear Fit" corresponds to the best linear fit (parameter value equal to -0.071), "20% less cure" uses a 20% lower value for the cure rate, and "5% higher HR" corresponds to a 20% higher value for the hazard ratio for baseline survival. Percentage of variation was chosen as the approximate minimum that increases/decreases prostate cancer mortality by 5%, relative to the original model fit.

Figure 4: Relationship between Cure Probability and lead-time.*



* Cure probability denotes the effect of screening. It is given by the expression, Cure probability = exp(cure parameter * lead-time). Cure parameter equals -0.22. Each point in the plot denotes a year.

Modelling Active Surveillance

Table 8: Modelling referral to treatment in AS*

Event	Modelling	Parameter(s)
Volume Progression	Indirectly Modelled, may occur if in absence of screening there would be an increase in T-stage.	Probability of volume progression, given an increase in T-stage.
Gleason Upgrade	Directly modelled, may occur if in absence of screening Gleason would increase.	Sensitivity for Gleason upgrade
Clinical Diagnosis $	Time of clinical detection in absence of screening.	Hazard of clinical diagnosis per stage.
Treatment in absence of progression	Randomly selected from all men in AS.	Probability of treatment in absence of evidence of progression or clinical diagnosis.

* The parameters of the natural history model are calibrated to ERSPC data (7). The parameters related to referral to treatment in AS are calibrated to JH-AS (11) cohort data, including the number of men experiencing volume progression or gleason upgrade, men treated without evidence of progression and the 5 year treatment free survival.

$ For the clinical diagnosis event, all the related parameters are calibrated to SEER data and denote an additional hazard of clinical detection in the US, relative to the European situation.

Table 9: Comparison of observed treatment free survival (TFS) and progression in JH-AS cohort with predicted by MISCAN.*

		Detection Probabilities at each AS round				
Detection Probabilities (Gleason/>T2a /≤T2a)	Observed	0/0/0	1/1/1	0.5/0.5/0.5	0.5/0.4/0.1	0.5/0.4/0.1
Time in AS	(a)	(b)	(c)	(d)	(e)	(f)
1		0.99	0.99	0.99	0.99	0.99
2	0.81	0.97	0.56	0.77	0.88	0.86
3		0.95	0.36	0.58	0.77	0.74
4		0.92	0.24	0.43	0.67	0.64
5	0.59	0.90	0.17	0.31	0.58	0.55
Grade Progression	106	0	127	105	111	109
Volume Progression	129	0	334	264	129	125
Treatment without Progression	67	69	34	41	44	66

* The 2011 follow-up report of the JH cohort (11) showed there were no prostate cancer deaths in the AS cohort. The latest update (5) showed a prostate cancer survival of 99.9% at 10 years. MISCAN predicted 2 prostate cancer deaths in against the observed 0 in (11).
a: As in Tosoian et al (11). This was updated in (12), where the TFS at 5 years was 61%.
b: if a man has volume or gleason progression it is not detected. Only clinically detected men are treated.
c: if a man has volume or gleason progression it is detected with 100% probability.
d: if a man has volume or gleason progression it is detected with 50% probability.
e: if a man has gleason progression it can be detected with 50% probability. If a man is in T2aGS6 stage or in > T2aGS6 , then the probability of detecting volume progression, is, respectively, 10% or 40%.

f: Similar to (d), but with a probability of leaving AS without progression equal to 0.04 at every AS round.

Table 10: Biopsy Compliance during Active Surveillance §

Year of Follow-up	Biopsy Compliance
1	81%*
2	73%
3	66%
4	60%*
5	58%
6	55%
7	53%*
8	45%
9	39%
10	33%*
>10	33%

§ All numbers marked with " * " are the observed biopsy compliances in the PRIAS cohort (13). The numbers in between are extrapolated using the compound annual growth rate formula. After 10 years in AS, we assume the biopsy compliance remains 33%.

Quality of Life: Calculation of Post-Treatment Utility

Table 11: Adverse symptom proportions and their utilities by treatment modality.*§

Symptom	Baseline	24 Months*	Baseline	24 Months*	Utility
Urinary Problem (Incontinence)	2	8	2	5	0.83
Bowel Problem	1	1	3	11	0.71
Sexual Problem	12	43	18	37	0.89
"No Problem" &	100	63	100	70	1

* All utilities come from (14), and the estimates of the proportion of symptoms at baseline and 24 months are based on (15).

& We assume that utility at baseline (that is, before cancer diagnosis) equals 1. The utility decrements are thus applied to the extra adverse events caused by radical treatment.

§ Average post-recovery treatment utility is calculated as the utility decrement due to symptoms times the approximate proportion of men who experience symptoms. For radical prostatectomy this is equal to,

(0.08-0.02)* 0.83 + (0.01-0.01)*0.71 + (0.43-0.12)*0.89 + 0.63*1 = 0.956,

And using a similar calculation for radiation therapy this equals 0.951. Since for simplification we assume there is an equal chance of getting either treatment, the average post-recovery treatment utility is 0.95.

References

1. Wever EM, Draisma G, Heijnsdijk EA, de Koning HJ. How does early detection by screening affect disease progression?: Modelling estimated benefits in prostate cancer screening. Med Decis Making. 2011;31:550-58.

2. Draisma G, Postma R, Schröder FH, et al. Gleason Score, age and screening: modelling dedifferentiation in prostate cancer. Int J Cancer. 2006;119:2366-71.

3. EAM Heijnsdijk, , TM de Carvalho, A Auvinen, *et al*.J Natl Cancer Inst. 2014;107(1):366.

4. Mariotto AB, Etzioni R, Krapcho M, et al. Reconstructing PSA Testing Patterns Between Black and White Men in the US from Medicare Claims and the National Health Interview Survey. Cancer. 2007;109(9):1877-86.

5. Wever EM, Draisma G, Heijnsdijk EA, et al. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. J Natl Cancer Inst. 2010 Mar 3;102(5):352-5.

6. Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer. 2012;118:5955-63.

7. 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.

8. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.

9. de Carvalho TM, Heijnsdijk EAM and de Koning HJ. Screening for Prostate Cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136:1600-7.

10. 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.

11. Tosoian JJ, Trock BJ, Landis P, et al. Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. J Clin Oncol 2011;29:2185-90.

12. Tosoian J, Mamawala M, Epstein J, et al. Intermediate and Longer-Term Outcomes From a Prospective Active-Surveillance Program for Favorable-Risk Prostate J Clin Oncol. 2015;33:3379-85.

13. 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.

14. Stewart ST, Lenert L, Bhatnagar V, et al. Utilities for prostate cancer health states in men aged 60 and older. Med Care. 2005;43:347-55.

15. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med. 2008;358:1250-61.

The End

Summary,
Acknowledgments
&
About the author

Summary

Introduction

Prostate cancer is the second most diagnosed cancer and the fifth deadliest cancer worldwide. The number of prostate cancer cases substantially increased in the US and Western Europe since the introduction of the PSA (prostate-specific antigen) test in the 90's. On the other hand, prostate cancer related mortality is experiencing a downward trend, which is attributed to improvements in treatment and early detection.

Two major randomized clinical trials evaluated whether prostate cancer screening reduces prostate cancer mortality. While one found no effect of screening on prostate cancer mortality (PLCO, in the US), the other found a 21% prostate cancer mortality reduction due to screening (ERSPC, in several European countries). Due to this, prostate cancer screening is considered to be controversial. Several reasons may explain this discrepancy in the outcomes between the two trials. Likely the most important explanation is the high proportion of men who were PSA tested in the control arm of the PLCO trial.

In general screening may cause both positive and negative effects. On the positive side, we consider prostate cancer mortality reduction and reduction in advanced disease. On the negative side, we consider overdiagnosis and overtreatment. Overdiagnosis consists on the diagnosis of disease through early detection, that would never become symptomatic. In case treatment is offered, this person is overtreated. In prostate cancer, this has a severe effect on the quality of life of men. First, an unnecessary diagnosis of cancer may cause major psychological harm. Secondly, treatment of prostate cancer may cause major side effects, like impotence, incontinence or bowel problems.

The goal of this thesis is, with the help of a simulation model, to predict the harms and benefits of several screening and treatment policies. We aim to find a policy which maximizes prostate cancer mortality reduction while keeping the harms of screening as low as possible.

Aim I: Screening

In this part, we aim to find the best possible prostate cancer screening policy. In chapter 2, we projected the results of more than 80 different screening policies, and we found that every screen above age 70 adds a substantial layer of overdiagnosis.

We can reduce overdiagnosis by 30% by stopping screening at age 70, and by more than 60% if we stop at age 66, compared to stopping screening at age 74. In chapter 3, we added costs and utilities, to each procedure associated with screening and treatment and we studied whether prostate cancer screening can be cost-effective. We found that using a cost-effectiveness threshold of 100,000 euros per quality-adjusted life year gained, a limited screening program (55-59, biannual) can be cost-effective.

In chapters 4 and 5, we moved away from a one-size-fits-all screening policy. In chapter 4 we used lifetables which differ per level of comorbidity to estimate the harms and benefits of screening. In chapter 5 we used lifetables and estimated screening patters for the African-American in order to find what causes the higher incidence of prostate cancer, among African-Americans, relative to the general US population. In both chapters 4 and 5, we found that cohort specific screening policies may be more efficient given the differential patterns of life expectancy and disease incidence of the different sub-populations.

Aim II: Active surveillance

In this aim, screen-detected low-risk men (and in some cases intermediate risk men) are referred to Active Surveillance, instead of immediate treatment. Active Surveillance for prostate cancer consists of the carefully monitoring of newly screen-detected men with, with PSA tests and/or repeat biopsies, resulting in referral to treatment if progression of the disease is observed. The aim of Active Surveillance is to keep the quality of life of men diagnosed with cancer, while avoiding the danger of missing the cure window, by progressing to advanced disease. As in Aim I, we projected the harms and benefits of screening combined with Active Surveillance. Two key questions that were evaluated throughout are: 1) Should intermediate risk men be included in AS? ; 2) How many biopsies should we do during AS?

In Chapter 6, we studied the effects of different Active Surveillance protocols, differing by biopsy frequency and whether or not we admit intermediate risk men on AS, on overtreatment and prostate cancer mortality. We found that including more patients in AS (from very low-risk, to low risk and intermediate risk), resulted in a higher reduction of overdiagnosis, though at a cost of a slightly increased prostate cancer mortality risk. On the other hand, it seems to be more effective to include only low-risk men and have a reduced number of biopsies (triannual) than to include low and intermediate risk men with annual biopsies. In Chapter 7 we stratified men by level of comorbidity at screen-detection (as in Chapter 3). While for people with high risk disease, there could be benefit in treatment even for men

with significant comorbidities, for men with low risk disease and no comorbidity, Active Surveillance should be the primary treatment option. For men with intermediate risk disease, level of comorbidity may be important, since younger men without comorbidity have a profile similar to high risk men.

In Chapter 8 we added quality of life and costs, and studied whether screening combined with AS is cost-effective. We found as in Chapter 3, that only limited screening, i.e. low frequency or early stopping age can be cost-effective. Additionally, we found that AS is more cost-effective than immediate treatment, independently of the intensity of screening. In Chapter 9 we studied when men should stop AS if no progression is detected. We compared Active Surveillance with Conservative Management, an approach where men are not treated unless the cancer becomes symptomatic. We found that for men older than 65, the benefit of AS compared to Conservative Management is limited. On the other hand, for men younger than 65, a relatively intensive biopsy protocol is recommended, with at least 7 biopsy rounds.

Aim III: Uncertainty in simulation models

All results in this thesis are subject to different forms of uncertainty, namely uncertainty due to simulation error, uncertainty around the parameter estimates and uncertainty due to assumptions used to build the simulation model. In this aim we investigate what is the effect of parameter uncertainty on the model outcomes. Our main goal is to obtain a distribution and confidence interval for model outcomes. This is a difficult task to perform with MISCAN, since many model runs are needed in order to build confidence interval. In order to make this task feasible we use a statistical technique called Gaussian process regression. We run MISCAN at a few selected input parameter configurations and collect the corresponding outcomes. Based on this data, we fitted a Gaussian process regression model that we used to obtain an estimate of the values of the outcomes for other parameter configurations of interest. We found an empirical confidence interval for overdiagnosis between 37% and 48%, with a 42% as the point estimate.

Conclusion

Screening is more effective when applied to younger age groups. Though some lives will be saved for older men, this comes at a cost of a large amount of overdiagnosis, and loss in quality of life. Active Surveillance is an effective way to reduce overtreatment. It should become the default option for low-risk men older

than 60. The differences in outcomes between the different AS protocols are small. Our findings support a less frequent biopsy schedule (triannual instead of annual), and inclusion of low-risk men only. Personalized screening and treatment strategies are needed in order to improve the harms and benefits trade-off of prostate cancer care, namely, besides age group, also race and comorbidity should be taken into account, and in the future, biomarkers, genetic markers and MRI-imaging may help to improve screening and treatment decisions.

Samenvatting

Introductie

Prostaatkanker is de tweede meest gediagnosticeerde kanker en de vijfde meest dodelijke kanker wereldwijd. Het aantal gevallen van prostaatkanker is aanzienlijk toegenomen in de VS en West-Europa sinds de invoering van de PSA (prostaat-specifiek antigeen) test in de jaren '90. Anderzijds is er een dalende trend te zien in prostaatkankersterfte, die wordt toegeschreven aan verbeteringen in de behandeling en aan vroege opsporing.

Twee grote gerandomiseerde studies hebben geëvalueerd of vroeg opsporing van prostaatkanker de prostaatkankersterfte vermindert. Terwijl de ene studie geen effect van screening op prostaatkankersterfte (PLCO, in de VS) heeft gevonden, vond de andere studie een 21% daling van prostaatkankersterfte als gevolg van screening (ERSPC, in verschillende Europese landen). Hierdoor wordt vroege opsporing van prostaat kanker als controversieel gezien. Verschillende redenen kunnen dit verschil in de resultaten verklaren. De belangrijkste verklaring is waarschijnlijk het hoge aandeel mannen in de controle-arm van de PLCO trial die een PSA test hebben ondergaan.

In het algemeen kan screening zowel positieve als negatieve effecten veroorzaken. Aan de positieve kant, prostaatkankersterfte reductie en een vermindering van vergevorderde ziekte. Aan de negatieve kant, overdiagnose en overbehandeling. Overdiagnose is een door vroege opsporing veroorzaakte diagnose van ziekte, die zonder de vroege opsporing nooit symptomatisch zou zijn geworden. In het geval dat de behandeling wordt aangeboden, wordt deze persoon overbehandeld. Bij prostaatkanker, heeft dit ernstige gevolgen op de kwaliteit van leven van mannen. Ten eerste, kan een onnodige diagnose van kanker grote psychische schade veroorzaken. Ten tweede, kan de behandeling van prostaatkanker ernstige

bijwerkingen, zoals impotentie, incontinentie of darmproblemen veroorzaken.

Het doel van dit proefschrift is, met de hulp van een simulatiemodel, de voor- en nadelen van verschillende screening en behandeling strategieën te voorspellen. Wij streven naar een beleid dat prostaatkankersterfte vermindert terwijl de nadelen van screening zo laag mogelijk blijven.

Deel I: Vroege Opsporing

In dit deel willen we de beste mogelijke vroege opsporingsbeleid vinden. In hoofdstuk 2, hebben we de resultaten van meer dan 80 verschillende strategieën geprojecteerd, en we vonden dat elke extra screen ronde boven de leeftijd van 70 een aanzienlijke hoeveelheid overdiagnose toevoegt. We kunnen overdiagnose verminderen met 30% door het stoppen met screening op leeftijd 70, en met meer dan 60% als we stoppen op leeftijd 66, in vergelijking met het stoppen van screening op leeftijd 74. In hoofdstuk 3 hebben we kosten en utiliteiten aan elke fase in het proces van screening en behandeling toegevoegd en we hebben onderzocht of vroege opsporing voor prostaatkanker kan kost-effectief zijn. Als we een drempel van € 100.000 per gewonnen voor kwaliteit van leven gecorrigeerde levensjaar gebruiken, vinden we dat een beperkt screeningsprogramma (55-59, tweejaarlijks) kosteneffectief kan zijn.

In hoofdstuk 4 gebruikten we lifetables die verschillen per comorbiditeit niveau, en we hebben onderzocht hoe de voor- en nadelen van screening ermee veranderen. In hoofdstuk 5 hebben we gebruik gemaakt van Afro-Amerikaanse specifieke lifetables en screening patronen om uit te vinden wat de oorzaak van de hogere incidentie van prostaatkanker, onder Afro-Amerikanen is, in verhouding tot de algemene Amerikaanse bevolking. In beide hoofdstukken 4 en 5, hebben we gevonden dat subpopulatie specifieke screening strategieën efficiënter kunnen zijn, gezien de verschillende patronen van levensverwachting en incidentie.

Deel II: Active Surveillance

In Deel II worden screen-gedetecteerde laag risico mannen (en in sommige gevallen intermediair risico mannen) verwezen naar Active Surveillance (AS), in plaats van onmiddellijke actieve behandeling. Active Surveillance voor prostaatkanker bestaat uit het zorgvuldig bewaken van de screen-gedetecteerde mannen met PSA testen en / of herhaalde biopten. Pas als progressie van de ziekte wordt bewezen volgt een verwijzing naar behandeling. Het doel van Active Surveillance is om de kwaliteit van leven van de gediagnosticeerd mannen te houden, terwijl ze het gevaar van

ongeneeslijk worden en/of ontwikkelen van gevorderde ziekte vermijden. Net als in Deel I, hebben we de voor- en nadelen van screening in combinatie met Active Surveillance gekwantificeerd. Twee belangrijke algemene vragen, in deze gedeelte zijn: 1) Kunnen intermediair risico mannen naar AS worden verwezen? ; 2) Hoeveel biopsieën moeten we doen tijdens AS?

In Hoofdstuk 6 onderzochten we de effecten van verschillende Active Surveillance protocollen, per biopsie frequentie en met en zonder inclusie van intermediair risico mannen, op overbehandeling en prostaatkanker sterfte. We vonden dat het verwijzen van meer patiënten naar AS (van zeer laag risico, laag risico en tot intermediair risico), resulteerde in een hogere reductie van overdiagnose, maar tegen een prijs van een licht verhoogd prostaatkankersterfterisico. Aan de andere kant lijkt efficiënter slechts lage risico mannen te includeren met een beperkt aantal biopsieën (driejaarlijkse) dan laag en gemiddeld risico mannen met jaarlijkse biopten te includeren. In hoofdstuk 7 stratificeren wij mannen per comorbiditeit niveau op de tijd van screen-detectie (zoals in hoofdstuk 3). Terwijl voor mensen met een hoog risico het voordelig is om te kiezen voor behandeling, zelfs voor mannen met comorbiditeiten, moet voor mannen met een laag risico ziekte en geen comorbiditeit, Active Surveillance de primaire behandelingsoptie zijn.
Voor patiënten met een intermediair risico, kan het niveau van comorbiditeit van belang zijn, omdat jongere mannen zonder comorbiditeit een profiel vergelijkbaar hebben als hoog risico mannen.

In Hoofdstuk 8 hebben we kwaliteit van leven en kosten toegevoegd, en hebben we onderzocht of screening in combinatie met AS kost-effectief kan zijn. Vergelijkbaar met hoofdstuk 3 hebben we gevonden dat een beperkte screeningsprogramma, dat is een programma met een lage screenfrequentie of vroege stop leeftijd, doelmatig kan zijn. Bovendien vonden we dat AS kosten-effectiever is dan onmiddellijke behandeling, onafhankelijk van de intensiteit van screening. In hoofdstuk 9 onderzochten we wanneer moeten mannen stoppen met AS als er geen voortgang in de ziekte wordt gedetecteerd. We vergeleken Active Surveillance met conservatieve behandeling, een aanpak waar mannen niet worden behandeld, tenzij de kanker symptomatisch wordt. We hebben gevonden dat bij mannen ouder dan 65, AS in weinig winst resulteert in vergelijking met conservatieve behandeling. Anderzijds kunnen we voor mannen jonger dan 65, een relatief intensieve biopsie protocol aanbevelen, inclusief ten minste 7 biopsie rondes.

Deel III: Onzekerheid in simulatie modellen

Alle resultaten in dit proefschrift zijn onderhevig aan verschillende vormen van onzekerheid, namelijk de onzekerheid als gevolg van simulatie error, de onzekerheid rond de parameter schattingen en de onzekerheid als gevolg van de aannames die gebruikt zijn om het simulatie model te ontwikkelen. In Deel III onderzoeken we wat het effect van de parameter onzekerheid is op de modeluitkomsten. Ons belangrijkste doel is om een verdeling en een betrouwbaarheidsinterval voor modeluitkomsten te schatten. Dit is een moeilijke taak om met MISCAN uit te voeren, aangezien veel model runs nodig zijn om het betrouwbaarheidsinterval te schatten. Om deze taak mogelijk te maken, gebruiken we een statistische techniek genaamd Gaussian Process Regression. We hebben MISCAN gerund bij een aantal geselecteerde input parameter configuraties en de bijbehorende resultaten verzameld. Op basis van deze gegevens, hebben we een Gaussian Process Regression model gebouwd en een schatting gemaakt van de waarden van de uitkomsten voor andere input parameter configuraties. We hebben een empirische betrouwbaarheidsinterval voor overdiagnose gevonden tussen 37% en 48%, met 42% als puntschatting.

Conclusie

Screening is effectiever wanneer het toegepast wordt op jongere leeftijdsgroepen. Hoewel ook bij oudere mannen sommige levens zullen worden gered, zal screening bij hen gepaard gaan met een grote hoeveelheid overdiagnose en verlies in kwaliteit van leven. Active Surveillance is een effectieve manier om overbehandeling te verminderen. Het moet de standaard behandelingsoptie worden voor laag-risico mannen ouder dan 60. We vinden dat de verschillen in uitkomsten tussen de verschillende AS protocollen klein zijn. Onze bevindingen ondersteunen een minder frequent biopsie schema (driejaarlijkse in plaats van jaarlijks), en AS aan te bieden aan mannen met een laag risico. Gepersonaliseerde screening en behandelingsstrategieën zijn nodig om de afweging tussen voor- en nadelen van prostaatkanker te verbeteren. Met name moet er naast de leeftijdsgroep, ook met ethniciteit en comorbiditeit rekening gehouden worden. In de toekomst kunnen biomarkers, genetische markers en MRI -Imaging helpen om beslissingen over screening en behandeling te verbeteren.

Acknowledgements

Many people contributed directly and/or indirectly to this book. In first place, I would like to thank my promotor, Prof. Harry de Koning for giving me this opportunity. Your critical eye and experience was essential, especially when writing my first paper. Eveline, as my daily supervisor, you always had an open door, for advice and support, it was a pleasure to work with you during this 4.5 years. Additionally, I would like to thank all members of the thesis commission, for taking their time to evaluate this thesis.

I would like to thank all fellow members or collaborators of CISNET-Prostate, namely, Ruth Etzioni, Roman Gulati, Alex Tsodikov and Sigrid Carlsson, among others: the CISNET meetings were always an outstanding source of new ideas and inspiration and greatly contributed to this book and my development as a researcher. I would also like to thank all participant members of the ERSPC meetings, and from the Urology department at Erasmus MC, for their helpful comments.

In my first months at Erasmus, I was fortunate to have Elizabeth as my own "MISCAN" guru. Your help was essential! During my final months at Erasmus I also took great pleasure in guiding Katerina, hope you enjoy your time at Erasmus MC! Additionally, Luc and Katerina: I am very thankful for your comments on the Gaussian Process article.

I was very lucky, to share my office room and corridor with very friendly and helpful colleagues (too many to name them all! : Frank, Else, Lea, Kevin, Suzette, Reinier, Amir, Maggie, Karen, Frederik, Ivana, Adi, David, Luc, Sandra, Carlijn, Branko, Esther, …), and I hope to keep in touch with some of you in the near future.

I would like to thank my mom, for all the support, and the multiple people that I met through Couchsurfing, in Rotterdam, in Lisbon and around the world which made these 4.5 years truly memorable!

About the author

Tiago Marques was born on the 15th February 1988, in Lisbon, Portugal. He completed his secondary school education at Escola Secundaria Antonio Gedeao, Almada. Afterwards he decided to study Economics at ISEG, University of Lisbon. During the last year of his studies he decided that it was time to go abroad, and switch to a more quantitative study.

In September 2008 he started to study Econometrics and Operations Research at the University of Groningen, first as a pre-Master, and then as a Master Student. In 2011, he moved to Rotterdam, for a job at the consulting company Pharmerit, where he also did his Master thesis.

In April 2012, he started at the Department of Public Health, Erasmus Medical Center as a Junior Researcher. During his time at the department he worked with a microsimulation model of prostate cancer screening. Most of his research was performed within the CISNET (Cancer Intervention and Surveillance Modelling Network) network, which is a National Cancer Institute (US) sponsored international consortium of modelling groups, for different cancer sites. Since December 2016, he works at the Department of Applied Health Research, University College London, continuing his research on prostate cancer screening.

The research findings on the harms and benefits of prostate cancer screening and Active Surveillance are presented in this thesis.

Publications and Working Papers

Published

1. **de Carvalho TM**, Heijnsdijk EA, de Koning HJ. When should Active Surveillance for prostate cancer stop if no progression is detected?. Prostate. 2017;77(9):962-969.

2. Tsodikov A, Gulati R, **de Carvalho TM**, Heijnsdijk EA, Hunter-Merrill RA, Mariotto AB et al. Is prostate cancer different in black men? Answers from three natural history models. Cancer. 2017. doi: 10.1002/cncr.30687

3. **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 2016. doi: 10.1111/bju.13542.

4. **de Carvalho TM**, Heijnsdijk EA, de Koning HJ. Estimating the individual benefit of immediate treatment or active surveillance for prostate cancer after screen-detection in older (65+) men. Int J Cancer 2016;138(10):2522-8.

5. Carlsson SV, **de Carvalho TM**, Roobol MJ, Hugosson J, Auvinen A, Kwiatkowski M, et al. Estimating the harms and benefits of prostate cancer screening as used in common practice versus recommended good practice: A microsimulation screening analysis. Cancer 2016. doi: 10.1002/cncr.30192.

6. **de Carvalho TM**, Heijnsdijk EA, de Koning HJ. Screening for prostate cancer in the US? Reduce the harms and keep the benefit. Int J Cancer 2015;136(7):1600-7.

7. Heijnsdijk EA, **de Carvalho TM**, Auvinen A, Zappa M, Nelen V, Kwiatkowski M., et al. Cost-effectiveness of prostate cancer screening: a simulation study based on ERSPC data. J Natl Cancer Inst 2014;107(1):366.

8. Lansdorp-Vogelaar I, Gulati R, Mariotto AB, Schechter CB, de **Carvalho TM**, Knudsen AB, et al. Personalizing age of cancer screening cessation based on comorbid conditions: model estimates of harms and benefits. Ann Intern Med 2014;161(2):104-12.

Working Papers

1. **de Carvalho TM**, Heijnsdijk EA, de Koning HJ. Cost-Effectiveness of Active Surveillance compared to Immediate Treatment: A modelling study. Submitted.

2. **de Carvalho TM**, Heijnsdijk EA, Coffeng L, de Koning HJ. Evaluating parameter uncertainty in a simulation model of cancer using emulators. Submitted.

3. **de Carvalho TM**, Heijnsdijk EA, de Koning HJ. Using a Gaussian process emulator to model many screening policies. Work in Progress.

PhD Portfolio

Activity	Period	ECTS
General Courses		
Survival Analyses	Summer 2012	1.9
Computational Econometrics	Fall 2012	0.7
Planning and Evaluation of Screening	Spring 2013	1.4
Introduction to Joint Modelling of Longitudinal and Survival Outcomes	Spring 2013	0.7
Principles of Genetic Epidemiology	Summer 2015	0.7
Case-Control Studies	Summer 2015	0.7
Causal Mediation Analyses	Summer 2015	0.7
Seminars and Workshops		
Seminars at the Department of Public Health, Erasmus MC Rotterdam	2012-2016	6.0
Methodology Club, Department of Public Health, Erasmus MC Rotterdam	2012-2016	1.0
Seminars at the Department of Biostatistics, Erasmus MC, Rotterdam	2012-2014	0.6
Seminars at the Department of Epidemiology, Erasmus MC, Rotterdam	2012-2016	0.6
Presentations		
Cancer Intervention and Surveillance Modelling Network (CISNET) meetings, National Cancer Institute, USA.	2012-2016	6.0
Methodology Club, Department of Public Health, Erasmus MC, Rotterdam.	2012-2016	1.8
Screen-Section Meeting, Department of Public Health, Erasmus MC, Rotterdam	2012-2016	1.8
European Randomized study of Screening for Prostate Cancer Meeting, Antwerpen and Gothemburg.	2013-2014	1.2
Cancer Control Joint Action (CanCon), Rotterdam	2016	0.6
International Conferences		
Active Surveillance for low-risk prostate cancer conference, European School of Oncology, Amsterdam, Milan.	2014, 2016	0.6
International Microsimulation Association (IMA) Conference , Esch/Belval, Luxembourg.	2015	0.6
Society for Medical Decision Making, Biennial European Conference, London.	2016	0.6
Teaching		
Biostatistics I Practicum, NIHES program, Erasmus Medical Center	Fall 2015	0.6
Review		
Reviewer for the Journal of the National Cancer Institute (JNCI)	2015	0.6
Reviewer for the British Journal of Cancer (Brit J Cancer)	2016	0.6
Total	2012-2016	30.0