Effects of Prostate Cancer Screening and Treatment

Elisabeth Wever

ISBN: 978-94-6169-296-2

Effects of Prostate Cancer Screening and Treatment
Thesis, Erasmus University

Statue on cover: © Kees Verkade, Oude Garnalenvisser (1969) c/o Pictoright Amsterdam 2012
Cover photo: René and Peter van der Krogt
Design, lay-out and print: Optima Grafische Communicatie, Rotterdam, The Netherlands

The studies reported in this thesis were primarily funded by the National Cancer Institute and the Netherlands Organisation of Health Research and Development (ZonMw). For any grant pertaining to specific studies the reader is referred to the individual papers published in their respective journals.

This thesis was printed with financial support of the J.E. Jurriaanse Stichting, the Department of Public Health Erasmus MC Rotterdam and the Erasmus University Rotterdam.

Effects of Prostate Cancer Screening and Treatment

Effecten van prostaatkanker
screening en behandeling

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 10 oktober 2012 om 11.30 uur

door
Elisabeth Marie Wever
geboren te Curaçao

ERASMUS UNIVERSITEIT ROTTERDAM

PROMOTIECOMMISSIE

Promotor:	Prof.dr. H.J. de Koning
Overige leden:	Prof.dr. C.H. Bangma
	Prof.dr. P.J.E. Bindels
	Prof.dr. P.M.M. Bossuyt

CONTENTS

Chapter 1

General Introduction

1.1 PROSTATE CANCER EPIDEMIOLOGY

Prostate cancer is the second most frequently diagnosed cancer of men worldwide. The number of new cases worldwide was estimated at 899,000 and accounted for 13.6% of all cancers in men in 2008. With an estimated 258,000 deaths in 2008, prostate cancer is the sixth leading cause of death from cancer in men (6.1% of the total).[1]

Prostate cancer is most common in Australia/New Zealand, Northern and Western Europe, and Northern America (Figure 1.1). The incidence of prostate cancer has been rising in these regions since the early nineties largely because of the widespread practice of PSA testing. Especially in the US the incidence increased rapidly after the introduction of prostate specific antigen (PSA) testing. While incidence also increased in the Netherlands, the increase was less dramatic than in the US (Figure 1.2).

Prostate cancer mortality rates have been decreasing since 1996 in both the Netherlands and the US (Figure 1.2). Possible explanations for this trend are improvements in prostate cancer treatment and screening for prostate cancer.

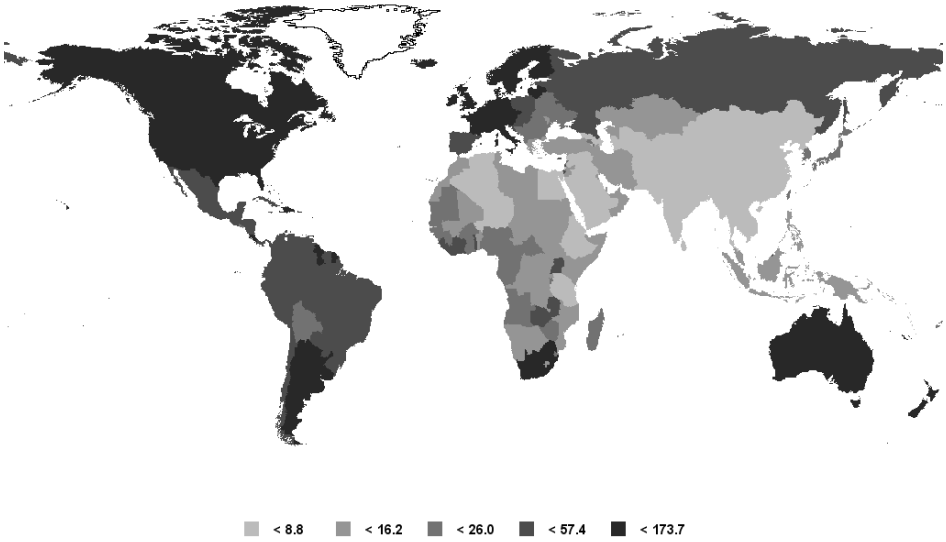


Figure 1.1: World age-standardized prostate cancer incidence by country (rate per 100,000 men). [Source: GLOBOCAN 2008][2]

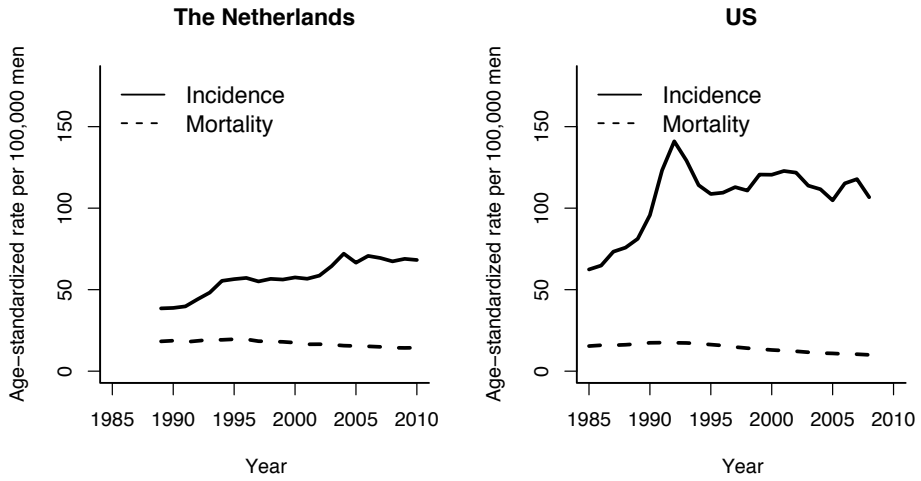


Figure 1.2: World age-standardized incidence and mortality rates of prostate cancer in the Netherlands[3] (1989-2010) and in the US[4] (1985-2008).

1.2 PROSTATE CANCER

Prostate cancer is cancer that develops in the prostate, a gland that is part of the male reproductive system. The prostate is a walnut-sized gland, located beneath the urine bladder, surrounding the proximal part of the urethra.

1.3 DIAGNOSTIC MODALITIES

Early prostate cancer causes no symptoms. It is only in its later stages that symptoms become evident. These may include blood in urine, difficulty urinating and bone pain.[5] There are several tests that can be used to determine if there is an increased risk for the presence of prostate cancer: a prostate specific antigen (PSA) test, a digital rectal examination (DRE) and transrectal ultrasonography (TRUS).

The PSA test is a blood test that measures the serum level of PSA in the blood. PSA is a protein produced by prostate cells. An increased PSA level indicates an increased prostate cancer risk.[6] However, an increased PSA level may also be due to other causes, such as benign prostatic hyperplasia (enlargement of the prostate) or prostatic inflammation.[7]

A DRE involves the palpation of the prostate through the rectal wall to determine the presence of irregularities, hardened areas or lumps. An abnormal DRE indicates an increased prostate cancer risk.[8] The sensitivity of the DRE is limited because of its inability to detect tumors deep within the prostate gland.

The prostate can also be examined for cancer using TRUS, a test that produces an image of the prostate, which may detect signs of prostate cancer.[9] TRUS has become indispensable for guiding biopsies of the prostate.
The results from a PSA-test, DRE or TRUS provide an indication about the risk of prostate cancer. The only way of diagnosing prostate cancer is by a histological examination of prostate tissue. The common method for obtaining prostate tissue for the examination is by a biopsy, where tissue cores are taken from the prostate using a needle.

1.4 STAGING

The extent of the disease is usually classified according the Tumor-Node-metastasis (TNM) classification (Table 1.1), which describes the anatomic extension of prostate cancer at different points in the disease process. Prostate cancer usually grows slowly and is initially confined to the prostate gland (clinical stage T1-T2), where it may not cause serious harm.

Table 1.1: Tumor, node, metastasis (TNM) classification of prostate cancer (2002 version[10]).

Primary tumor (T)
TX: Primary tumor cannot be assessed
T0: No evidence of primary tumor
T1: Clinically unapparent tumor neither palpable nor visible by imaging
T1a: Tumor incidental histological finding in 5% or less of tissue resected
T1b: Tumor incidental histological finding in more than 5% of tissue resected
T1c: Tumor identified by needle biopsy (for example, because of elevated PSA)
T2: Tumor confined within prostate
T2a: Tumor involves one-half of one lobe or less
T2b: Tumor involves more than one-half of one lobe but not both lobes
T2c: Tumor involves both lobes
T3: Tumor extends through the prostate capsule
T3a: Extracapsular extension (unilateral or bilateral)
T3b: Tumor invades seminal vesicle(s)
T4: Tumor is fixed or invades adjacent structures other than seminal vesicles, such as external sphincter, rectum, bladder, levator muscles, and/or pelvic wall
Regional Lymph Nodes (N)
NX: Regional lymph nodes were not assessed
N0: No regional lymph node metastasis
N1: Metastasis in regional lymph node(s)
Distant Metastasis (M)
MX: Distant metastasis were not assessed
M0: No distant metastasis
M1: Distant metastasis
M1a: Non-regional lymph node(s)
M1b: Bone(s)
M1c: Other site(s) with or without bone disease

However, after this stage the cancer may grow outside the prostatic capsule (clinical stage T3), which may be followed by spread of the cancer into surrounding organs (clinical stage T4). Eventually, prostate cancer cells might spread into the lymph nodes (N1) or other parts of the body (M1), for example into the bones.

1.5 GRADING

The grade of the tumor expresses the degree of abnormality of the tissue and thus the aggressiveness of the tumor. The Gleason grading system[11] is the most commonly used method for grading prostate cancer. In the Gleason grading system growth patterns are divided into five categories with scores 1 (well differentiated) to 5 (poorly differentiated) (Figure 1.3). The Gleason score is obtained by summing the scores of the most common and the second most common growth pattern observed. Gleason scores range therefore from 2 to 10. The Gleason score is an important prognostic factor for a patients' disease specific outcome.[12-14]

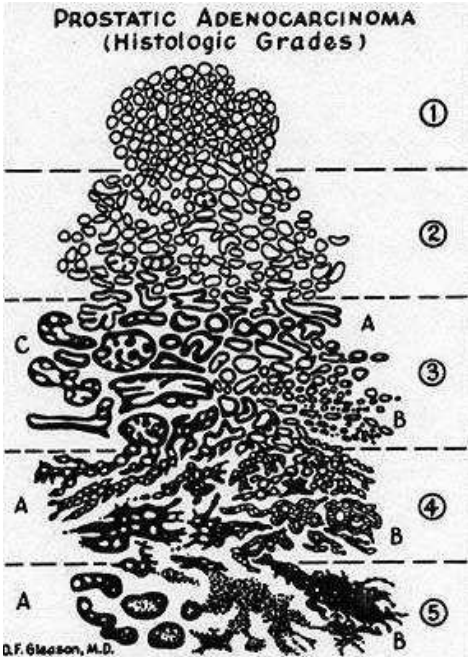


Figure 1.3: Appearance of cancer cells, ranked according to the Gleason grade, adapted from[15]

1.6 TREATMENTS

Men with localized prostate cancer have the option to go into an expectant management program or to receive one of the curative treatments. The two expectant management approaches are active surveillance and watchful waiting. Active surveillance is an approach,

where the patients will be regularly monitored with a variety of tests and if over time the disease appears to be advancing curative treatment might be performed. The aim of active surveillance is to delay or avoid treatment of prostate cancer and the potential adverse side-effects. Watchful waiting is an approach where treatment will not be given or given unless symptoms appear or change, and is usually an alternative for men in whom curative treatment is not an option due to age or co-morbidity. Watchful waiting is different from the active surveillance approach, in which deferred treatment has a curative intent.

The most common curative treatments can be categorized in two groups: radical prostatectomy and radiation therapy. Radical prostatectomy involves removing the complete prostate gland and the seminal vesicles. Radiation therapy involves radiation of the prostate gland and minimizing exposure to surroundings.[5]

Curative treatment can have a favorable effect on the prostate cancer specific mortality. The randomized Scandinavian Prostate Cancer Group Study Number 4 (SPCG-4)[16] showed a decrease in the cumulative incidence of death (6% in absolute terms) from prostate cancer, when comparing radical prostatectomy with watchful waiting in clinically-detected localized prostate cancer. There are no randomized control trial results showing the effectiveness of radiation therapy. However, some comparative effectiveness studies showed a lower benefit for radiation therapy over radical prostatectomy.[17-18]

However, all curative treatments can result in side-effects. Radical prostatectomy often leads to urinary and sexual problems. It has been observed that at 52 months after surgery 31% of men had urinary leakage problems and 88% erectile dysfunction, while only 12% had urinary leakage problems and 31% erectile dysfunction before the surgery. Radiation therapy mainly leads to sexual and bowel problems. For men receiving radiation therapy it has been observed that at 52 months, 64% of men had erectile dysfunction and 11% bowel bother, while only 40% had erectile dysfunction and 3% bowel bother before radiation therapy.[19]

Metastasized prostate cancer can no longer be cured. However, temporary suppression of the disease is possible using different forms of hormonal therapy, chemotherapy or castration.

1.7 SCREENING

Screening involves testing for cancer (or for conditions that may lead to cancer) in people who have no cancer-specific symptoms. Screening can help finding cancers at an early stage, when there may be a better chance of curing the cancer. The currently most promising procedure to screen for prostate cancer is a PSA test, followed by a biopsy test if the PSA level is higher than a specific threshold. A threshold level of 3 or 4 ng/ml is being used in ongoing randomized control trials to evaluate the efficacy of prostate cancer screening.

Benefits

The benefit that can be achieved by screening and subsequent early treatment is a reduced risk of advanced disease, a reduced risk of dying from prostate cancer and an increased life-expectancy. To evaluate the efficacy of prostate cancer screening, two large randomized trials were initiated in the early 1990s: the European Randomized Study of Screening for Prostate Cancer (ERSPC) in Europe and the Prostate, Lung, Colorectal, and Ovary (PLCO) trial in the United States. The results of the ERSPC trial, at a median follow-up of eleven years, showed a significant prostate cancer mortality reduction in the screening group of 21%, and of 29% in men actually screened.[20-21] After a median follow-up of 14 years, the Göteborg trial, one center of the ERSPC trial, showed a 44% prostate cancer mortality reduction in the screening group and a 56% reduction for men screened at least once.[22] The PLCO trial found no mortality reduction in the screening group, however the rate of screening in the control arm (contamination) was high and the attendance rate (compliance) at biopsy low.[23-24]

Harms

PSA screening is however also associated with considerable unfavorable effects. An adverse effect of screening is overdiagnosis, i.e. the detection of cancers that would not have been diagnosed during the patients' lifetime if they had not been screened. Furthermore, all men with screen-detected prostate cancer have to live more years with the knowledge that they have prostate cancer. Men with screen-detected prostate cancer who opt for curative treatment risk living many years with the side-effects of treatment, which would otherwise be symptom-free years.[19]

The over-diagnosis rate and mean lead time (the period by which the diagnosis is advanced due to screening) are large because prostate cancer is a slow growing cancer, and because it is especially common in older men. So, although many men may harbor prostate cancer, they often will not experience the negative effects of the cancer because they may die of other causes before the cancer has a chance to progress.

Over-diagnosis rates and lead times are measures that can not directly be observed and can only be estimated by using models simulating the natural history of cancer. For men screened in the ERSPC Rotterdam an over-diagnosis rate of 66% and a mean lead time of 7.9 years have been estimated.[25] However, several estimates have been published for the over-diagnosis rate (23-66%)[25-29] and mean lead time (5-12 years).[26, 30-32] The explanation for these wide ranges of estimates is often that over-diagnosis rates and mean lead times are dependent on definition, calculation methods and study population, including the amount of diagnosis already in the population.[25]

PSA screening dissemination in US and the Netherlands

In the US, PSA testing was approved in 1986 by the Food and Drug Administration for monitoring prostate cancer progression. Despite the lack of evidence of screening efficacy from

randomized trials at that time, the test was rapidly adopted for screening. By the year 2005, 54% of men within ages 50–84 years have had at least one PSA test. The frequency of a first PSA test and of repeat tests are illustrated in Figure 1.4.[33] The incidence of prostate cancer in the US increased simultaneously with the increase of first PSA tests (Figure 1.2 and Figure 1.4).

In Europe, PSA testing has not been as widespread as in the US.[34-36] Evidence from surveys in ERSPC centers suggests however an increasing rate over time. For example, in Rotterdam in the Netherlands, yearly PSA testing increased from 3.5% (1996-1998) to 5.7% (1997-2000).[34] Also, according to a self-report study of the Statistics Netherlands (CBS) in 2001 14.4% of men reported that they have had at least one PSA test in the previous five years, while in 2009 this percentage was 24.8%.[37]

Currently, most major U.S. medical organizations[38-41] and the European Association of Urology[42] recommend that clinicians discuss the potential benefits and adverse effects of PSA screening with their clients, consider their clients' preferences, and individualize screening decisions. However, the U.S. Preventive Services Task Force (USPSTF) recommends against screening for prostate cancer for all ages.[43]

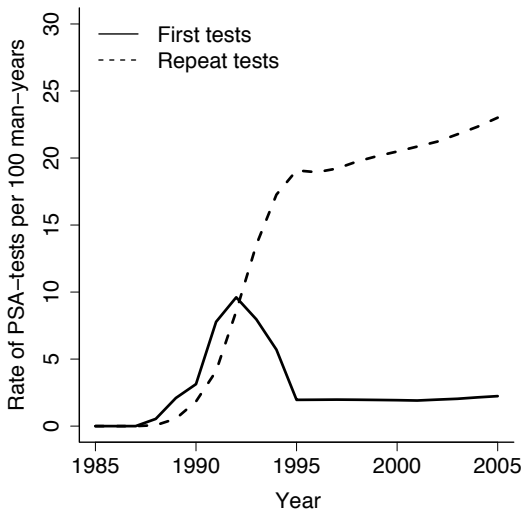


Figure 1.4: Frequency of first PSA tests and repeat tests in the US population as based results of an updated version of the model of Mariotto et al.[33] The frequencies are for men aged 50-84.

1.8 MODELING CANCER PROGRESSION, SCREENING AND TREATMENT

Randomized control trials are essential for determining the benefits and harms of a screening procedure or a treatment. However, because of the long follow-up that is needed, it is often difficult to determine the lifetime risk of the benefits and harms with only the empirical data. Models that are validated on trial data can be used to predict these lifetime risks. Also, models

can be used to extrapolate trial results to different screening settings, for example different screening ages and intervals. Therefore, models have been important tools for estimating benefits and harms of interventions for early detection, for instance the screening of breast cancer,[44-46] cervical cancer[47-48] and colorectal cancer.[49-50]

Models can also be used for understanding the impact of behavioral changes, screening or treatment on observed cancer incidence and mortality. For example, Moolgavkar et al.[51] quantified with six models (including the Miscan-lung model) the impact of changes in smoking behaviours which started in the mid-1950s on lung cancer mortality in the United States. They showed that approximately 795,851 US lung cancer deaths were averted during the period 1975–2000 and that these numbers represent approximately 32% of lung cancer deaths that could have potentially been averted during the period. Also, Berry et al.[52] showed with seven models (including the Miscan-breast model) that mammography screening and treatment have helped equally to reduce the rate of death from breast cancer in the United States.

1.9 CISNET

The Cancer Intervention and Surveillance Modeling Network (CISNET) is a consortium of research groups sponsored by the US National Cancer Institute (NCI). In the CISNET-project, investigators use statistical modeling to estimate the impact of cancer control interventions in prevention, screening, and treatment.[53] Investigators with different models work cooperatively to explore differences between models in a systematic way. While each research group has areas of individual focus, in joint collaborations we try to understand differences and similarities across models. The idea behind this cooperation is that by working together, models are improved and modeling work becomes more transparent.

There are currently five cancer sites considered in the CISNET: breast, colorectal, esophagus, lung and prostate. MISCAN is the only non-American model involved in the CISNET-project and is involved in all cancer sites. Part of the work in this thesis was conducted in the prostate cancer CISNET-project.

1.10 ERSPC TRIAL

The European Randomized Study of Screening for Prostate Cancer (ERSPC) was initiated in the early 1990s to evaluate the effect of PSA screening on prostate cancer mortality.[54] In seven countries, 162,243 men were randomly assigned either to the intervention or the control arm. The screening interval was 4 years, with the exception of Sweden (2 years). Most centers used a PSA cutoff value of 3.0 ng/ml as an indication for biopsy, whereas others used

4.0 ng/ml, with additional tests for values between 2.5 and 4.0. Biopsies were administered by the screening centre at no charge to the subject, and reminders for biopsy appointments were sent if necessary. Treatment was performed according to local policies and guidelines.[55] The results of the ERSPC trial, at a median follow-up of 11 years, showed a significant prostate cancer mortality reduction of 29% in men screened.[21]

Part of the work in this thesis was conducted as part of the ERSPC trial. For all analyses we used ERSPC trial data for estimating the parameters of the progression of prostate cancer.

1.11 PLCO TRIAL

The prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial in the US was initiated in the early 1990s to determine whether screening for prostate cancer with PSA testing and DRE results in a reduction in prostate cancer mortality. A total of 76,685 men, aged 55–74 years, were enrolled at 10 screening centers and randomly assigned to the intervention (organized screening of annual PSA testing for 6 years and annual DRE for 4 years) or control (usual care, which included opportunistic screening) arms. Men with a positive PSA test were referred to their personal physician for follow-up. After 13 years of follow-up, there was no evidence of a mortality benefit for organized annual screening in the PLCO trial compared with opportunistic screening.[23]

The results of the two trials seem to be in conflict. However, there were many differences in the design and implementation of the two trials. Modeling can be used to quantify the impact of these differences.

1.12 MISCAN PROSTATE CANCER MODEL

In this thesis, the MISCAN prostate cancer microsimulation model has been used to evaluate the effects of screening and treatment on prostate cancer incidence and mortality. The MISCAN microsimulation model was developed at the Department of Public Health, at Erasmus Medical Center, the Netherlands, and has been used for modeling breast, prostate, cervix, colon and lung cancer screening.[26, 48, 56-58] A detailed description of the model and the data sources that informed its quantification can be found in the model appendix of this thesis, in previous studies[25-26, 59] and also in a standardized model profile.[60] In brief, MISCAN prostate cancer model is a micro-simulation program, which simulates the development of prostate cancer in individuals as a sequence of tumor states. In Figure 1.5 the transition through the different possible states is illustrated. The individual life-histories are first simulated in the absence of screening and subsequently with the changes that would occur when screening took place. First, the year of birth and age at death from other causes is simulated per

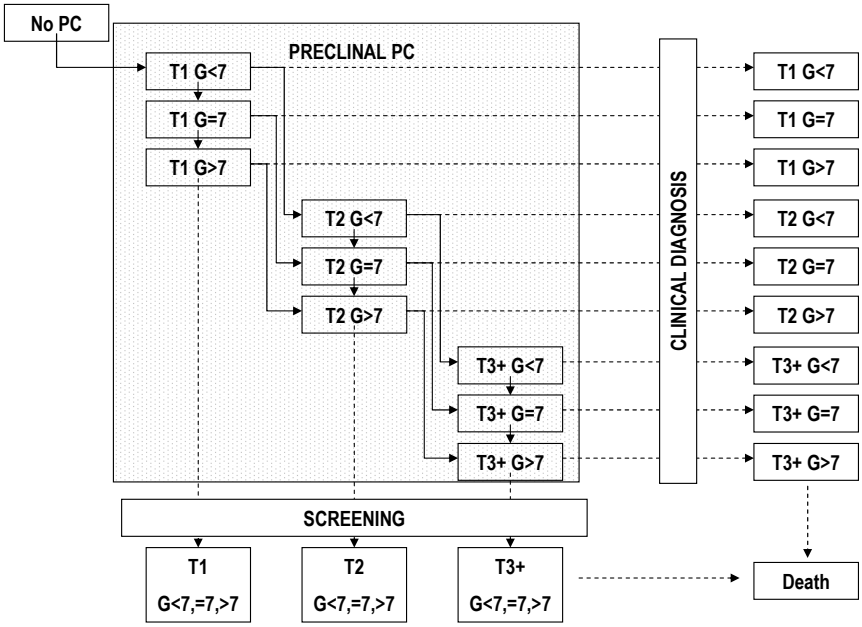


Figure 1.5: The MISCAN prostate cancer model. Prostate cancer develops from no prostate cancer via one or more screen-detectable preclinical stages to a clinically diagnosed cancer. In each preclinical state, a tumor may grow to the next clinical T-stage (T1, T2, or T3), dedifferentiate to a higher Gleason score (G < 7, G = 7, or G > 7), or become clinically diagnosed. There is also a risk that a local-regional tumor (M0) will develop into distant disease (M1). For simplicity, this possibility is not illustrated. Screening may detect cancers earlier in one of the preclinical screen-detectable states.

individual. Second, the progression of prostate cancer in the absence of screening is simulated. Prostate cancer may develop from no prostate cancer to a clinically diagnosed cancer through one or more screen-detectable preclinical stages. In each preclinical state, a tumor may grow to the next clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; and T3+, palpable, with extensions beyond the prostatic capsule), differentiated to a higher Gleason score (well differentiated, Gleason score 2–6; moderately differentiated, Gleason score 7; and poorly differentiated, Gleason score 8–10), progress from local-regional disease (M0) to distant disease (M1), or give rise to symptoms and become clinically diagnosed. The survival after diagnosis depends on the age, the stage of the cancer and the treatment received. Third, depending on the frequency and sensitivity of the screening test, preclinical cancers may be detected by screening. If detected by screening the cancer has a reduced risk of being fatal. By simulating life-histories with and without screening and/or treatment, the effects of screening and treatment can be assessed.

1.13 RESEARCH QUESTIONS AND OUTLINE OF THIS THESIS

In this thesis we analyzed the effect of prostate cancer screening and treatment. Part 1 covers studies that provide important quantitative information for ongoing debates about the effects of prostate cancer screening and treatment on observed incidence and mortality in the US population and PLCO trial. Additionally there is a study on the validation of the model by modeling observed prostate cancer mortality. Part 2 covers studies that present clinical outcomes which are necessary for making decisions about screening and treatment.

The following research questions will be addressed:

Part 1: Modeling observed prostate cancer incidence and mortality

- Can changes in primary treatment plausibly explain the observed decline in prostate cancer mortality by the year 2005 in the US? (Chapter 2)
- Is the PSA screening performance for detecting prostate cancers in the US population the same as in the ERSPC Rotterdam? (Chapter 3)
- Can we apply the commonly used assumption on the effect of screening on survival, the stage-shift assumption, to predict the observed prostate cancer mortality reduction in the ERSPC? (Chapter 4)
- What is the impact of PLCO control arm contamination on perceived PSA screening efficacy? (Chapter 5)

Part 2: Predicting clinical outcomes of prostate cancer progression, screening and treatment.

- What are the risks of clinical progression events for PSA-detected localized prostate cases who do not receive curative treatment? (Chapter 6)
- How do the lives of men who decide to be screened differ from the lives of men who decide not to be screened? And therefore, what is the anticipated loss in quality of life after a prostate cancer diagnosis and treatment that would be acceptable to decide in favor of screening? (Chapter 7)
- What are the quality of life effects of prostate cancer screening, and what is the number needed to screen? (Chapter 8)
- What are the benefits and harms of immediate treatment of PSA-detected local-regional prostate cancer, given the prognostic factors Gleason score, clinical T-stage and age? And for which combination of prognostic factors is immediate treatment most or least favorable? (Chapter 9)

Chapter 10 concludes this thesis with summary answers to and further discussion of the above research questions and directions for future research.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* Jun 17 2010.
2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin D. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 2010; http://globocan.iarc.fr. Accessed December 28, 2011.
3. Integraal Kankercentrum Registratie. http://cijfersoverkanker.nl/ (accessed February 20, 2012).
4. Surveillance, Epidemiology and End Results (SEER). http://seer.cancer.gov/ (accessed February 21, 2012).
5. Dollinger M, Rosenbaum EH, Cable G. *Everyone's Guide to Cancer Therapy* 3rd ed. Toronto: Andrews Mcmeel Publishing; 1997.
6. Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine.* Oct 8 1987;317(15):909-916.
7. Nadler RB, Humphrey PA, Smith DS, Catalona WJ, Ratliff TL. Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels. *J Urol.* Aug 1995;154(2 Pt 1):407-413.
8. Gosselaar C, Roobol MJ, Roemeling S, Schroder FH. The role of the digital rectal examination in subsequent screening visits in the European randomized study of screening for prostate cancer (ERSPC), Rotterdam. *European Urology.* Sep 2008;54(3):581-588.
9. Gosselaar C, Roobol MJ, Roemeling S, Wolters T, van Leenders GJ, Schroder FH. The value of an additional hypoechoic lesion-directed biopsy core for detecting prostate cancer. *BJU Int.* Mar 2008;101(6):685-690.
10. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A. *AJCC Cancer Staging Manual*. 7th ed. New York: Springer; 2010.
11. Gleason DF. Classification of prostatic carcinomas. *Cancer chemotherapy reports.* Mar 1966;50(3):125-128.
12. Adami HO. The prostate cancer pseudo-epidemic. *Acta oncologica (Stockholm, Sweden).* Apr 2010;49(3):298-304.
13. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
14. Reese AC, Cooperberg MR, Carroll PR. Minimal impact of clinical stage on prostate cancer prognosis among contemporary patients with clinically localized disease. *J Urol.* Jul 2010;184(1):114-119.
15. http://www.prostatelab.com/grading.htm.
16. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.
17. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. *Cancer.* Jul 1 2011;117(13):2883-2891.
18. Cooperberg MR, Vickers AJ, Broering JM, Carroll PR. Comparative risk-adjusted mortality outcomes after primary surgery, radiotherapy, or androgen-deprivation therapy for localized prostate cancer. *Cancer.* Nov 15 2010;116(22):5226-5234.
19. 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.* Aug 20 2005;116(2):291-296.
20. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
21. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine.* Mar 15 2012;366(11):981-990.
22. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *The lancet oncology.* Aug 2010;11(8):725-732.
23. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst.* Jan 18 2012;104(2):125-132.
24. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
25. 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.* Mar 18 2009;101(6):374-383.
26. 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.* Jun 18 2003;95(12):868-878.

27. Etzioni R, Penson DF, Legler JM, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst.* Jul 3 2002;94(13):981-990.
28. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics.* Mar 2008;64(1):10-19.
29. Zappa M, Ciatto S, Bonardi R, Mazzotta A. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. *Ann Oncol.* Dec 1998;9(12):1297-1300.
30. Auvinen A, Maattanen L, Stenman UH, et al. Lead time in prostate cancer screening (Finland). *Cancer Causes Control.* Apr 2002;13(3):279-285.
31. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *Jama.* Jan 25 1995;273(4):289-294.
32. Hugosson J, Aus G, Becker C, et al. Would prostate cancer detected by screening with prostate-specific antigen develop into clinical cancer if left undiagnosed? A comparison of two population-based studies in Sweden. *BJU Int.* Jun 2000;85(9):1078-1084.
33. 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.* May 1 2007;109(9):1877-1886.
34. Ciatto S, Zappa M, Villers A, Paez A, Otto SJ, Auvinen A. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. *BJU Int.* Dec 2003;92 Suppl 2:97-100.
35. Melia J, Moss S. Survey of the rate of PSA testing in general practice. *British Journal of Cancer.* Sep 1 2001;85(5):656-657.
36. Skovlund E, Fossa SD. [PSA analyses in three Norwegian counties in 1999]. *Tidsskr Nor Laegeforen.* Sep 10 2000;120(21):2539-2541.
37. CBS. Centrale Bureau van Statistieken. http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=03799&D1=291&D2=0&D3=0&D4=1-9&HD=120201-1423&HDR=T&STB=G1,G2,G3&CHARTTYPE=1 (accessed February 01, 2012). 2010.
38. Screening for prostate cancer. American College of Physicians. *Annals of Internal Medicine.* Mar 15 1997;126(6):480-484.
39. Lim LS, Sherin K. Screening for prostate cancer in U.S. men ACPM position statement on preventive practice. *American journal of preventive medicine.* Feb 2008;34(2):164-170.
40. Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW. Cancer screening in the United States, 2010: a review of current American Cancer Society guidelines and issues in cancer screening. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):99-119.
41. Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):70-98.
42. Heidenreich A, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. *European Urology.* Jan 2011;59(1):61-71.
43. Moyer VA. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine.* May 21 2012.
44. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. *Eur J Cancer.* Jan 2009;45(1):127-138.
45. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine.* Nov 17 2009;151(10):738-747.
46. van Ravesteyn NT, Miglioretti DL, Stout NK, et al. Tipping the Balance of Benefits and Harms to Favor Screening Mammography Starting at Age 40 Years: A Comparative Modeling Study of Risk. *Annals of Internal Medicine.* May 1 2012;156(9):609-617.
47. Mandelblatt JS, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. *JAMA.* May 8 2002;287(18):2372-2381.
48. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *J Natl Cancer Inst.* Feb 6 2002;94(3):193-204.
49. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Boer R, Wilschut J, Habbema JD. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. *Int J Cancer.* Mar 1 2009;124(5):1161-1168.
50. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine.* Nov 4 2008;149(9):659-669.

51. Moolgavkar SH, Holford TR, Levy DT, et al. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975-2000. *J Natl Cancer Inst.* Apr 4 2012;104(7):541-548.
52. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine.* Oct 27 2005;353(17):1784-1792.
53. National Cancer Institute, Cancer Intervention an Surveillance Modeling Network. http://cisnet.cancer.gov/. (Accessed on January 18, 2012).
54. Roobol MJ, Schroder FH. European Randomized study of Screening for Prostate Cancer (ERSPC): rationale, structure and preliminary results. *BJU Int.* Dec 2003;92 Suppl 2:1-122.
55. Wolters T, Roobol MJ, Steyerberg EW, et al. The effect of study arm on prostate cancer treatment in the large screening trial ERSPC. *Int J Cancer.* May 15 2010;126(10):2387-2393.
56. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Computers and biomedical research, an international journal.* Feb 1999;32(1):13-33.
57. Shi L, Tian H, McCarthy WJ, Berman B, Wu S, Boer R. Exploring the uncertainties of early detection results: model-based interpretation of mayo lung project. *BMC cancer.* 2011;11:92.
58. Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. *Journal of the National Cancer Institute.* 2006(36):56-65.
59. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
60. Wever EM, Heijnsdijk EA, Draisma G, de Koning HJ. Model profile of the MISCAN prostate cancer model. Department of Public Health, Erasmus MC, 2010. Accesed at: https://cisnet.flexkb.net/mp/pub/CISNET_ModelProfile_PROSTATE_ERASMUS_001_12152009_69754.pdf.

Part 1

Modeling observed prostate cancer incidence and mortality

Chapter 2

The prostate cancer conundrum revisited: Treatment changes and prostate cancer mortality declines

Ruth Etzioni, Roman Gulati, Alex Tsodikov, Elisabeth M. Wever, David F. Penson, Eveline A.M. Heijnsdijk, Jeffrey Katcher, Gerrit Draisma, Eric J. Feuer, Harry J. de Koning and Angela B. Mariotto

Cancer. Epub ahead of print in May 2012. Doi: 10.1002/cncr.27594.
Reprinted with permission from John Wiley & Sons.

ABSTRACT

Background

Prostate cancer mortality rates in the US declined by over 40% between 1991 and 2005. The impact of changes in primary treatment and adjuvant and neoadjuvant hormonal therapy on this decline is unknown.

Methods

Application of three independently developed models of prostate cancer natural history and disease detection under common assumptions about treatment patterns, treatment efficacy, and survival in the population. Primary treatment patterns are from the Surveillance, Epidemiology and End Results registry and hormonal therapy frequencies are from the CaPSURE database; treatment efficacies are based on estimates from randomized trials and comparative effectiveness studies of treatment alternatives. The models project prostate cancer mortality without PSA screening and in the presence and absence of treatment benefit. Impact of primary treatment is expressed as a fraction of the difference between observed mortality and projected mortality in the absence of treatment benefit.

Results

The three models project that changes in treatment explain 22–33% of the mortality decline by 2005. These contributions are accounted for mostly by surgery and radiation therapy, which increased in frequency until the 1990s; hormonal therapies contributed little to the mortality decline by 2005. Assuming that treatment benefit is less for older men, changes in treatment explain only 16–23% of the mortality decline by 2005.

Conclusions

Changes in primary treatment explain a minority of the observed decline in prostate cancer mortality. The remainder of the decline is likely due to other interventions, such as PSA screening and advances in the treatment of recurrent and progressive disease.

INTRODUCTION

Since the early 1990s we have witnessed a spectacular decline in prostate cancer mortality in the US. Between 1991 and 2005 alone, prostate cancer mortality declined by 42% from 103 to 60 deaths per 100,000 men aged 50–84 y. This remarkable success story coincided with dramatic changes in the control of the disease: the widespread adoption of prostate-specific antigen (PSA) screening beginning around 1987, advances in treatment of early stage tumors, and changes in the detection and treatment of recurrent and progressive disease.

Because of the simultaneous dissemination of PSA screening and changes in treatment, a clear explanation for the drop in prostate cancer deaths has been elusive. In a 2003 editorial titled "The Prostate Cancer Conundrum," Albertsen questioned the relative roles of primary surgery and adjuvant hormonal therapy for localized disease in explaining the mortality trends.[1] Rates of surgery surged in the 1980s following the development of nerve-sparing techniques for radical prostatectomy. Randomized trial results indicate that radical prostatectomy improves disease-specific survival relative to watchful waiting, with a 38% reduction in the risk of prostate cancer death.[2] Hormonal therapy, previously reserved for men with advanced cancers, is particularly efficacious when used in combination with external-beam radiation therapy, and its use dramatically increased in the mid- to late 1990s.[3]

The role of PSA screening in explaining the drop in disease-specific deaths has also been questioned[4] but has not been conclusively determined. Long-awaited results from two large prostate cancer screening trials failed to convincingly establish screening benefit, with the European trial showing a 20% lower disease-specific mortality rate in the screening arm over a median of nine years[5] and the US trial showing no difference between the control and screening arms after seven years of complete follow-up.[6] However, it is generally recognized that since men on the control arm of the US trial received "usual care," which included routine screening,[7] the results should be interpreted as a comparison between moderate and high screening intensities.[8]

The Cancer Intervention and Surveillance Modeling Network (CISNET) prostate group was formed to quantify the relative contributions of screening and treatment changes to the mortality declines. Previously, CISNET prostate models were used to show that early detection due to screening could account for approximately 45–70% of the decline in prostate cancer mortality under a "stage-shift" mechanism for screening benefit. The stage-shift mechanism specifies that disease shifted to an earlier stage by screening enjoys a corresponding improvement in disease-specific survival. This mechanism is a central motivator underlying all cancer screening studies; however, the extent to which it holds is not known conclusively in the case of prostate cancer.

In this article, we take a different approach and quantify the fraction of the mortality decline plausibly due to treatment changes among men with non-metastatic disease. To do this, we model the dissemination and benefits of first-line treatment (radical prostatectomy and

radiation therapy alone or in combination with hormonal therapy) and project their impact on mortality in the absence of screening. The results are informative about the likely role of treatment changes in explaining prostate cancer mortality declines. In addition, they are suggestive of a potential role for screening and/or other practice changes, such as treatment for recurrent or progressive cancer.

METHODS

The CISNET paradigm

The CISNET approach is, at its core, a model of disease natural history, representing the individual experience of disease onset and progression, diagnosis, and death in the absence of any interventions of interest. Interventions, such as screening and/or treatment, are then superimposed based on analyses of patterns of care in the population and on known efficacy from randomized trials or assumed mechanisms of benefit.

In the present setting the models first produce projections of prostate cancer mortality in the absence of screening and treatment among cases diagnosed from 1975 because limited population-representative data are available before 1975 to inform the natural history models. By "absence of treatment" we mean in the absence of treatment benefit, as all projections under this setting assume that primary treatment interventions are not beneficial (i.e., the hazard ratio for disease-specific survival equals 1.0 relative to conservative management). The models also project mortality in the presence of treatment but in the absence of screening, i.e., assuming that stage and grade distributions at diagnosis would have remained as observed in the pre-PSA era. We project mortality in the absence of screening because projections in the presence of screening would rely on an assumed survival benefit of screening, a benefit with greater uncertainty than the benefits of primary treatments.

The impact of treatment occurs through changes in treatment distributions (Figure 2.1) as well as through treatment benefit (treatment-specific hazard ratios for disease-specific survival that are less than 1.0 relative to conservative management). We use the terms "in the presence of treatment" and "in the presence of changes in treatment" interchangeably. By comparing the mortality projections in the presence and absence of treatment with observed disease-specific mortality trends, we can quantify the fraction of the mortality decline associated with treatment. For example, if a projection in the presence of treatment lies half way between our projection in the absence of treatment and observed mortality, we would conclude that treatment alone (due to treatment benefit and changes in treatment patterns) accounts for approximately 50% of the observed mortality decline.

The CISNET prostate working group consists of three groups developing independent models of prostate cancer natural history informed by common information on patterns of screening, disease incidence, and other-cause mortality. Each natural history model is differ-

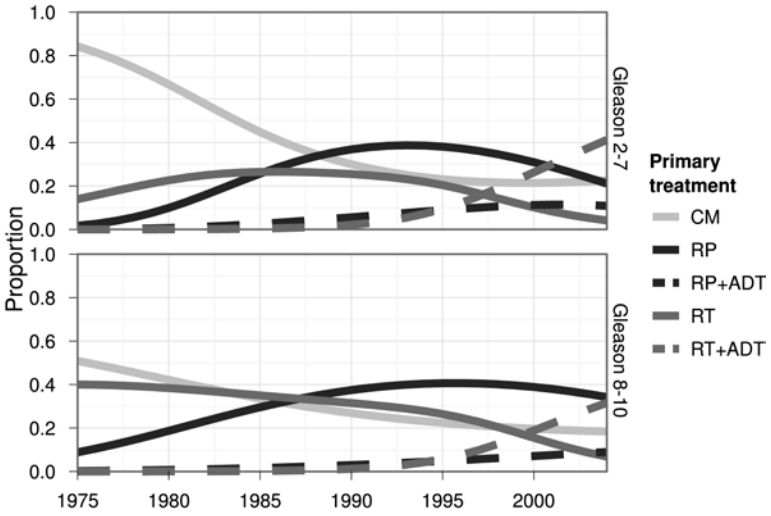


Figure 2.1: Primary treatment dissemination patterns in the United States are illustrated. Radical prostatectomy (RP) is defined by Surveillance, Epidemiology, and End Results (SEER) codes 50, 58, 60, and 68 before 1997 and by SEER codes 50 and 70 beginning in 1998. Radiation therapy (RT) is defined according to SEER categories as beam radiation, radioactive implants, radioisotopes, a combination of beam with implants or isotopes, or radiation with method or source unspecified. Androgen-deprivation therapy (ADT) data are from CaPSURE (Cancer of the Prostate Strategic Urologic Research Endeavor). CM indicates conservative management.

ent, but the models are each calibrated using age-, year-, stage-, and grade-specific prostate cancer incidence from the Surveillance, Epidemiology, and End Results (SEER) program before and after the introduction of PSA screening. The calibrated natural history models are then combined with common information on treatment patterns and disease-specific survival to project prostate cancer mortality under plausible assumptions about treatment efficacy.

In the next section we briefly describe each natural history model and the calibration methods used. We then detail survival modeling procedures, our data sources, and assumptions regarding treatment efficacy.

Model structures

Detailed descriptions of individual models and a joint report comparing the models are available at http://cisnet.cancer.gov/prostate/profiles.html.

The FHCRC model

In the Fred Hutchinson Cancer Research Center (FHCRC) model, the risk of disease onset, and associated Gleason grade category which is fixed at onset, depends on age. Disease progresses from localized to metastatic stages and from latent to symptomatic states based on risks that depend on grade-specific PSA levels. Distributions of PSA growth rates were estimated using longitudinal PSA measurements from men in the control arm of the Prostate Cancer

Prevention Trial.[9] Given individual PSA trajectories and natural histories, PSA screening patterns,[10] biopsy compliance frequencies observed in the US-based Prostate, Lung, Colorectal, and Ovarian cancer screening trial,[11-12] and trends in biopsy sensitivity in the population,[13] risks of transitioning from one state to the next were estimated using maximum likelihood to obtain parameter estimates that best reproduce SEER incidence.[14]

The MISCAN model

In the Erasmus University Medical Center MIcrosimulation SCreening ANalysis (MISCAN) prostate model, cancer development is modeled as a semi-Markov process governing transitions from one state to the next. In addition to the healthy state, there are 18 states in the natural history of prostate cancer that are derived from combinations of clinical T (T1, T2, and T3) and M (M0 and M1) stages in the TNM staging system and Gleason grade (well, moderately, and poorly differentiated). Cancers in each state may be clinically diagnosed or detected by a PSA test and subsequent biopsy, the probability of which is combined in a single sensitivity parameter that is state-specific. Model parameters (progression rates between states and test sensitivities) were estimated using data from the Rotterdam section of the ERSPC.[15-16] For calibration to the US situation, we re-estimated the test sensitivity parameters and estimated an additional stage-specific risk of clinical diagnosis to capture different pre-PSA disease diagnosis patterns in the US as compared with Europe. US-specific estimates for the parameters were obtained by calibrating the model to the observed age-specific incidence and age-specific metastatic stage distribution using maximum likelihood.[17]

The UMICH model

The University of Michigan (UMICH) natural history model consists of disease-free, preclinical, and clinical states. An analytic formulation first estimates age- and year-specific disease incidence based on PSA screening patterns,[10] assuming parametric distributions for age at onset and for time from onset to diagnosis, and increasing test sensitivity with time since onset. As in the MISCAN model, test sensitivity reflects both the diagnostic properties of the test itself and the frequency and sensitivity of any subsequent biopsy. Parameters are estimated by averaging over these distributions and calibrating the resulting marginal incidence against observed incidence.[18]

Next, disease stage (SEER local-regional or distant) and grade category (Gleason score 2–7 or 8–10) at diagnosis are estimated based on time from onset to diagnosis and mode of detection (screen or clinical) via a multinomial logistic model.[19] Maximum likelihood estimation of the joint model of age-specific incidence trends and stage/grade distributions informs the distributions of these clinical characteristics.

Modeling survival

In the absence of treatment, all three models generate disease-specific survival based on SEER data among men diagnosed just prior to the PSA era, during the calendar interval 1983–1986. A Poisson regression model is fit to the disease-specific survival frequencies, censoring deaths due to other causes and adjusting for age, stage, and grade at diagnosis and initial treatment (radical prostatectomy, radiation therapy, both, or neither). Then the fitted survival curve for men receiving neither treatment is used to predict disease-specific survival times under no screening and no initial therapy. Relative to local-regional survival, trends in survival for distant stage disease have remained fairly constant over time.[20]

Treatment dissemination and efficacy

Dissemination of treatment is modeled based on two data sources. Trends in primary treatments—radical prostatectomy (RP) and radiation therapy (RT)—are based on data from SEER, which records first cancer-directed therapy received. Trends in receipt of adjuvant or neoadjuvant androgen deprivation therapy (ADT), also called hormonal therapy, are based on the Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) database.[3, 21] CaPSURE was initiated in 1995 to document community trends in prostate cancer practice patterns, epidemiology, and outcomes. It is a longitudinal, observational database accruing data from 40 urologic practice sites over its history. There are currently over 14,000 men enrolled in CaPSURE. CaPSURE collects approximately 1,000 clinical and patient-reported variables. Clinical information is collected by the treating urologist at baseline and with each follow-up visit. Figure 2.1 documents trends in primary treatment by Gleason category.

We model five initial treatment courses for local-regional disease: conservative management (CM), RP, RP+ADT, RT, and RT+ADT. These treatments are modeled because they represent the predominant treatment interventions used to treat prostate cancer, because their utilization has changed over time, and because there is quantitative evidence regarding efficacy available from randomized trials. We grouped three-dimensional conformal beam external beam radiotherapy, intensity modulated radiation therapy, and low- and high-dose interstitial brachytherapy into a single RT. This was based on a recent comparative effectiveness review from the Agency for Healthcare Research and Quality's Evidence-Based Practice Center at Tufts University which found no studies reporting a significant difference in overall survival or biochemical failure among the various forms of radiation.[22] Unfortunately, there are no randomized comparisons of all treatments; consequently, we integrate evidence from several sources.

A brief summary of comparative effectiveness results for these treatments is as follows. There is evidence that RP is more efficacious than CM with a relative risk of 0.62.[2] There are no clinical trials directly comparing RP with RT; however, comparative effectiveness studies adjusting for case mix and progression risk consistently show a benefit for RP over RT with two recent studies producing adjusted relative risks of 0.45 and 0.47 for the endpoint of

disease-specific mortality.[23-24] There is evidence from clinical trials that RT+ADT is more efficacious than RT alone[25-27] and evidence from a recent comparative effectiveness study that RT+ADT is similar to RP (relative risk = 1.14 for RT+ADT relative to RP, p = 0.61[24]). For a recent review see Wilt et al.[28]

Based on these studies, we assume a hazard ratio of 0.62 for RP relative to CM and for RP+ADT relative to CM and apply this to prostate cancer-specific survival for untreated cases. However, to make the relative benefit of RT alone consistent with published studies would either require RT alone to be almost without benefit or RT+ADT to be far superior than RP. Therefore our assumed benefit range for RT compromises, reflecting lower benefit than either RP or RT+ADT, but not so low as to make RT completely ineffective. We also assume a time-varying relative risk associated with RT relative to CM to reflect the improvement in the efficacy of RT as more intense dose-delivery regimens evolved. Specifically, we assume the hazard ratio for RT relative to CM improved linearly from 0.9 in 1990 to 0.7 or 0.8 in 1995 and remained constant thereafter. This assumption implies a relative risk for RP versus RT that is either $0.62/0.8 \approx 0.77$ or $0.62/0.7 \approx 0.89$ after 1995. To reflect an even greater relative benefit of RP relative to RT alone, consistent with recent comparative effectiveness studies,[23-24] we also conduct a high-efficacy sensitivity experiment in which we use the more efficacious assumption for RT (hazard ratio 0.7 relative to CM after 1995) and lower the hazard ratio for RP relative to CM to 0.4, implying a relative risk for RP versus RT of $0.4/0.7 \approx 0.57$ after 1995.

We also consider age-specific hazard ratios for all curative treatments, based on the finding from the Scandinavian trial that RP was more beneficial in younger than in older men. Specifically, we consider hazard ratios for RP, RP+ADT, and RT+ADT relative to CM of 0.49 for men aged 50–64 y at diagnosis and 0.83 for men aged 65–84 y at diagnosis.[2] Corresponding hazard ratios for RT among men aged 50–64 y at diagnosis improve from 0.9 in 1990 to $0.49 \times 0.70/0.62 \approx 0.55$ and to $0.49 \times 0.80/0.62 \approx 0.63$ in 1995 remaining constant thereafter, while the hazard ratio for RT is constant at 0.9 for men aged 65–84 y at diagnosis for all years. In other words, we preserve the benefits of RT relative to RP within each age group. These four basic assumption sets (not including the high-efficacy sensitivity experiment) are summarized in Table 2.1.

Finally, we consider two additional sensitivity experiments, both using the more efficacious assumptions for RP (hazard ratio 0.62 relative to CM for all ages) and RT (hazard ratio 0.7 relative to CM after 1995). The first uses the UMICH model and allows prostate cancer incidence to continue its pre-PSA increase in the absence of screening, lowering the both distant stage incidence and prostate cancer mortality. The second uses the FHCRC model and assumes that all cases reported as receiving conservative management in SEER actually received radiation therapy as an extreme correction for possible underreporting of radiation therapy in SEER registries.[29]

Table 2.1: Assumed hazard ratios for prostate cancer survival for primary and hormonal treatments relative to conservative management.

Efficacy assumptions	RP, RP+ADT, and RT+ADT		RT by 1995	
	50–64 y	65–84 y	50–64 y	65–84 y
Assumption set 1	0.62	0.62	0.70	0.70
Assumption set 2	0.62	0.62	0.80	0.80
Assumption set 3	0.49	0.83	0.55	0.90
Assumption set 4	0.49	0.83	0.63	0.90

Notes: RP is radical prostatectomy, RT is radiation therapy, ADT is androgen deprivation therapy. All-age and age-specific hazard ratios for RP, RP+ADT, and RT+ADT are from Bill-Axelson et al.[2] with similarity between RP and RT+ADT based on Cooperberg et al.; [24] we assume RP+RT is similar to RP. The all-age hazard ratio for RT is 0.90 until 1990, improves linearly to 0.70 (Assumption set 1) or 0.80 (Assumption set 2) in 1995, and remains constant thereafter. The age-specific hazard ratio for RT for men aged 50–64 y at diagnosis is 0.90 until 1990, improves linearly to 0.55 (Assumption set 3) or 0.63 (Assumption set 4) in 1995, and remains constant thereafter; the hazard ratio for RT for men aged 65–84 y at diagnosis is a constant 0.90 across calendar years.

RESULTS

Age-adjusted prostate cancer incidence trends observed in the SEER registries are shown by stage in Figure 2.2. Also shown are corresponding incidence trends projected by the three models. The model projections replicate key features of the trends in local-regional incidence, including the rapid escalation in the late 1980s, the peak and initial decline in the early 1990s, and the stabilization at a higher level in the late 1990s. There is greater variability across models in the distant stage trends, though all models reproduce the scale of pre-PSA incidence and the rapid decline in the mid- to late 1990s.

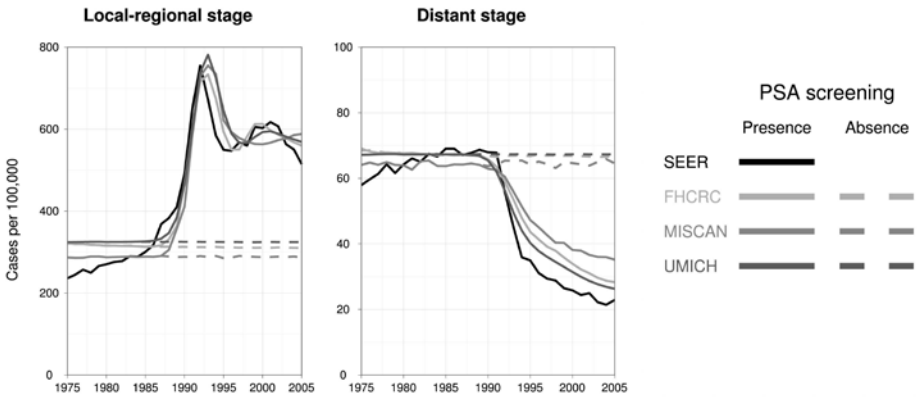


Figure 2.2: Age-adjusted SEER (black) and CISNET model-projected prostate cancer incidence in the presence (solid grays) and absence (dashed grays) of PSA screening.

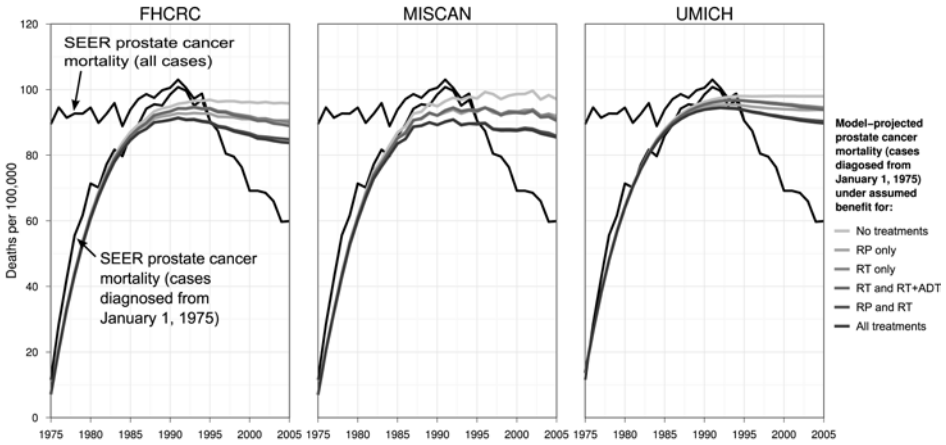


Figure 2.3: Age-adjusted SEER (black line) and CISNET model-projected (gray lines) prostate cancer mortality among cases diagnosed from January 1, 1975, under no primary treatment benefit or a combination of primary treatments. Model projections are based on assumption set 1. For comparison, the figure also shows SEER prostate cancer mortality among all cases.

We present incidence projections in the presence of PSA screening to show how well the calibrated natural history models perform relative to observed incidence. Based on these calibrated natural history models, under common assumptions about treatment and survival, we project prostate cancer mortality in the absence of PSA screening.

Age-adjusted mortality projections allowing benefit for each treatment based on assumption set 1 are presented in Figure 2.3. All models reproduce the accumulation of prostate cancer deaths by 1985 and slightly underestimate the peak in 1991. All models project a more-or less constant continuation of mortality in the absence of treatment benefit and a modest decrease in the presence of treatment benefit. By 2005, mortality projections allowing benefit for all treatments represent up to one-third of the difference between mortality projected in the absence of treatment benefit and observed mortality.

Table 2.2 provides a quantitative summary of the estimated contribution of each treatment to the observed mortality decline. The three models generally agree that changes in RP and RT each played a role in the mortality decline. Under assumption set 1, RP explains 11–14%, RT explains 9–16%, and ADT explains 1–3% of the mortality decline relative to projected mortality in the absence of treatment benefit. Impacts are smaller when treatment is less beneficial for older men; under assumption set 4, corresponding impacts are 10–12%, 5–7%, and 1–3%. Changes in treatment explain from 22–33% to 16–23% across the four sets of assumed efficacy levels.

Despite conceptual differences across models about how prostate cancer develops and progresses, the models provide consistent results concerning the contributions of treatment to the difference between observed mortality in the year 2005 and the mortality that would have been expected in the absence of advances in treatment. Specifically, the models project

Table 2.2: Projected prostate cancer mortality rates per 100,000 men aged 50–84 y in 2005 by model, assumption about primary and hormonal treatment benefit, and treatment allowing benefit. The corresponding mortality rate from SEER was 59.9 in 2005.

Model	None	RP	RT	RT+ADT	RP+RT	RP+RT+ADT
FHCRC						
Set 1	95.8	90.6 (14.4)	90.0 (16.2)	88.9 (19.0)	84.8 (30.5)	83.8 (33.4)
Set 2	95.8	90.6 (14.4)	91.9 (10.7)	89.5 (17.4)	86.8 (25.1)	84.4 (31.8)
Set 3	95.8	91.3 (12.4)	92.8 (8.2)	91.8 (10.9)	88.4 (20.7)	87.4 (23.4)
Set 4	95.8	91.3 (12.4)	93.1 (7.4)	92.0 (10.5)	88.7 (19.8)	87.5 (23.0)
MISCAN						
Set 1	97.1	92.0 (13.7)	90.9 (16.6)	90.5 (17.7)	85.8 (30.3)	85.4 (31.3)
Set 2	97.1	92.0 (13.7)	93.0 (10.9)	91.5 (15.0)	88.0 (24.6)	86.4 (28.7)
Set 3	97.1	94.4 (7.3)	94.2 (7.8)	93.8 (8.8)	91.5 (15.1)	91.1 (16.2)
Set 4	97.1	94.4 (7.3)	94.4 (7.3)	93.9 (8.7)	91.6 (14.7)	91.1 (16.0)
UMICH						
Set 1	98.0	93.8 (10.8)	94.5 (9.1)	93.8 (10.9)	90.4 (20.0)	89.7 (21.6)
Set 2	98.0	93.8 (10.8)	95.6 (6.3)	94.3 (9.6)	91.5 (17.0)	90.1 (20.6)
Set 3	98.0	94.1 (10.1)	95.9 (5.3)	95.4 (6.8)	92.0 (15.6)	91.5 (16.9)
Set 4	98.0	94.1 (10.1)	96.1 (4.9)	95.4 (6.7)	92.2 (15.1)	91.6 (16.7)

Notes: Percent declines relative to the difference between mortality projected under no treatment benefit and SEER observed mortality are shown in parentheses. RP is radical prostatectomy, RT is radiation therapy, ADT is androgen deprivation therapy.

that prostate cancer mortality would have stabilized at just under 100 deaths per 100,000 men aged 50–84 y in the absence of treatment benefit. Our computation of the percent of the mortality decline explained by treatment trends in the year 2005 indicates a significant role for primary treatment, with treatment alone explaining up to one-third of the difference between the observed mortality rate in the year 2005 and the rate projected in the absence of treatment benefit.

Under our high-efficacy sensitivity experiment, changes in treatment still explained only about half (range across the three models 42–53%) of the decline in mortality by 2005. Allowing prostate cancer incidence to continue its pre-PSA increase in the absence of screening, the UMICH model projects that changes in treatment explained 30% rather than 22% of the decline in mortality by 2005. And assuming all cases reported as receiving conservative management in SEER actually received radiation therapy, the FHCRC model projects that changes in treatment explained 46% rather than 33% of the decline in mortality by 2005.

Thus, we conclude that advances in primary treatment likely played an important role in the dramatic drop in prostate cancer mortality observed since the early 1990s. However, changes in primary treatment alone do not explain the majority of the mortality decline.

DISCUSSION

The decline in prostate cancer mortality that began in the early 1990s has been striking and sustained. Between 1994 and 2005, prostate cancer deaths dropped by an average rate of 4.1% per year and they are still declining.

The present study uses comparative modeling to investigate one of the most plausible explanations for the mortality decline, i.e., changes in primary treatment, with the goal of shedding light also on the potential roles of screening and other interventions. Our results indicate that treatment explains a non-trivial fraction of the drop in disease-specific deaths, but the majority of the decline is likely explained by other factors such as screening or improvements in disease management after primary therapy. For example, with almost all patients being monitored with PSA after diagnosis, metastatic or potentially metastatic tumors are being re-treated considerably earlier.[30] Salvage treatments given at biochemical failure have been associated with significant improvements in disease-specific survival.[31] These changes in secondary disease management may have been primarily responsible for the early decline in mortality; based on recent screening trial results, we would not expect to see a substantial decline in mortality as early as was observed due to screening alone.

Our results rest on several key assumptions. First, each natural history model makes different assumptions about disease onset, progression, and diagnosis in the absence of screening. As a result, the three models project three estimates for the fraction of the mortality decline explained by treatment. We find that our conclusions are robust even given this inter-model uncertainty. Second, all models assume that disease incidence would have remained constant at pre-PSA levels after 1987. A sensitivity experiment found that our conclusions are robust even if disease incidence would have continued its increasing trend. Third, all models assume that baseline (in the absence of screening or treatment) prostate cancer survival remained constant in the PSA era. Even if this survival improved over time, perhaps due to advances in treating recurrent disease, this would have little impact on our results because it would imply similar relative differences between projected mortality rates in the presence and absence of changes in treatment.

Our study uses data from a variety of sources which are subject to limitations. Although SEER is the most authoritative resource for information on disease incidence and survival in the US, we note again that estimates of prostate cancer survival in the absence of screening are not available in the PSA era. We also use SEER data on the first course of cancer-directed therapy to estimate the frequencies of radical prostatectomy and radiation therapy. A sensitivity experiment found that our conclusions are robust even if all cases recorded as receiving conservative management in SEER actually received radiation therapy. Finally, our treatment efficacy estimates, which are based on the most rigorous and up-to-date results from randomized trials and comparative effectiveness studies, are still subject to moderate uncertainty. Our sensitivity experiments found that our conclusions are robust even assuming that

radical prostatectomy primarily benefits younger men and/or assuming that improvements in radiation technology achieved efficacy similar to radical prostatectomy.

The models use estimates of the efficacy of radical prostatectomy relative to conservative management from the benchmark Scandinavian randomized controlled trial.[2] A recent observational study[32] compared Medicare patients in the US who did and did not receive radical prostatectomy. After adjusting for selection, the study found an advantage for surgery, even among older men, who were not found to benefit significantly in the Scandinavian trial. If surgery is more efficacious in the US, then our results may be somewhat conservative since changes in the frequency of radical prostatectomy are associated with an important portion of the decline in mortality associated with primary treatment.

In conclusion, the results of this modeling study clearly identify a role for primary treatment changes in US prostate cancer mortality declines, but a large fraction of the decline is left unexplained. This clearly suggests a role for PSA screening, but it also indicates that we should not assume that screening is as effective as suggested by the overall drop in prostate cancer mortality observed in the PSA era. Indeed, there is a clear role for primary treatment change and possibly advances in treatment for recurrent or progressive disease, and there may be a synergy with earlier detection due to screening. Further modeling studies will investigate the extent to which screening and treatment jointly explain the mortality decline and will also highlight the role of other interventions such as advances in disease management for recurrent and metastatic disease.

REFERENCES

1. Albertsen PC. The prostate cancer conundrum. *J Natl Cancer Inst.* Jul 2 2003;95(13):930-931.
2. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.
3. Cooperberg MR, Grossfeld GD, Lubeck DP, Carroll PR. National practice patterns and time trends in androgen ablation for localized prostate cancer. *J Natl Cancer Inst.* Jul 2 2003;95(13):981-989.
4. Hankey BF, Feuer EJ, Clegg LX, et al. Cancer surveillance series: interpreting trends in prostate cancer--part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J Natl Cancer Inst.* Jun 16 1999;91(12):1017-1024.
5. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
6. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
7. Pinsky PF, Blacka A, Kramer BS, Miller A, Prorok PC, Berg C. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *Clin Trials.* Aug 2010;7(4):303-311.
8. Berg CD. The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial: the prostate cancer screening results in context. *Acta oncologica (Stockholm, Sweden).* Jun 2011;50 Suppl 1:12-17.
9. Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst.* Apr 19 2006;98(8):529-534.
10. 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.* May 1 2007;109(9):1877-1886.
11. Grubb RL, 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int.* Dec 2008;102(11):1524-1530.
12. Pinsky PF, Crawford ED, Kramer BS, et al. Repeat prostate biopsy in the prostate, lung, colorectal and ovarian cancer screening trial. *BJU Int.* Apr 2007;99(4):775-779.
13. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control.* Mar 2008;19(2):175-181.
14. 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.* Oct 2010;11(4):707-719.
15. 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.* Jun 18 2003;95(12):868-878.
16. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
17. 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.* Mar 3 2010;102(5):352-355.
18. Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Stat Med.* Aug 30 2006;25(16):2846-2866.
19. Chefo S, Tsodikov A. Stage-specific cancer incidence: an artificially mixed multinomial logit model. *Stat Med.* Jul 10 2009;28(15):2054-2076.
20. Cetin K, Beebe-Dimmer JL, Fryzek JP, Markus R, Carducci MA. Recent time trends in the epidemiology of stage IV prostate cancer in the United States: analysis of data from the Surveillance, Epidemiology, and End Results Program. *Urology.* Jun 2010;75(6):1396-1404.
21. Lubeck DP, Litwin MS, Henning JM, et al. The CaPSURE database: a methodology for clinical practice and research in prostate cancer. CaPSURE Research Panel. Cancer of the Prostate Strategic Urologic Research Endeavor. *Urology.* Nov 1996;48(5):773-777.
22. Bannuru RR, Dvorak T, Obadan N, et al. Comparative evaluation of radiation treatments for clinically localized prostate cancer: an updated systematic review. *Annals of Internal Medicine.* Aug 2 2011;155(3):171-178.
23. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. *Cancer.* Jul 1 2011;117(13):2883-2891.

24. Cooperberg MR, Vickers AJ, Broering JM, Carroll PR. Comparative risk-adjusted mortality outcomes after primary surgery, radiotherapy, or androgen-deprivation therapy for localized prostate cancer. *Cancer.* Nov 15 2010;116(22):5226-5234.
25. Bolla M, Collette L, Blank L, et al. Long-term results with immediate androgen suppression and external irradiation in patients with locally advanced prostate cancer (an EORTC study): a phase III randomised trial. *Lancet.* Jul 13 2002;360(9327):103-106.
26. D'Amico AV, Chen MH, Roehl KA, Catalona WJ. Preoperative PSA velocity and the risk of death from prostate cancer after radical prostatectomy. *New England Journal of Medicine.* Jul 8 2004;351(2):125-135.
27. Denham JW, Steigler A, Lamb DS, et al. Short-term androgen deprivation and radiotherapy for locally advanced prostate cancer: results from the Trans-Tasman Radiation Oncology Group 96.01 randomised controlled trial. *The lancet oncology.* Nov 2005;6(11):841-850.
28. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of Internal Medicine.* Mar 18 2008;148(6):435-448.
29. Jagsi R, Abrahamse P, Hawley ST, Graff JJ, Hamilton AS, Katz SJ. Underascertainment of radiotherapy receipt in Surveillance, Epidemiology, and End Results registry data. *Cancer.* Jan 15 2012;118(2):333-341.
30. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. *Jama.* May 5 1999;281(17):1591-1597.
31. Trock BJ, Han M, Freedland SJ, et al. Prostate cancer-specific survival following salvage radiotherapy vs observation in men with biochemical recurrence after radical prostatectomy. *Jama.* Jun 18 2008;299(23):2760-2769.
32. Hadley J, Yabroff KR, Barrett MJ, Penson DF, Saigal CS, Potosky AL. Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. *J Natl Cancer Inst.* Dec 1 2010;102(23):1780-1793.

Chapter 3

Prostate-specific antigen screening in the United States versus in the European Randomized Study of Screening for Prostate Cancer–Rotterdam

Elisabeth M. Wever, Gerrit Draisma, Eveline A.M. Heijnsdijk, Monique J. Roobol, Rob Boer, Suzie J. Otto and Harry J. de Koning

Journal of National Cancer Institute. 2010;102(5):352-355.
Reprinted with permission from Oxford University Press.

ABSTRACT

Dissemination of prostate-specific antigen (PSA) testing in the United States coincided with an increasing incidence of prostate cancer, a shift to earlier stage disease at diagnosis, and decreasing prostate cancer mortality. We compared PSA screening performance with respect to prostate cancer detection in the US population versus in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). We developed a simulation model for prostate cancer and PSA screening for ERSPC–Rotterdam. This model was then adapted to the US population by replacing demography parameters with US-specific ones and the screening protocol with the frequency of PSA tests in the US population. We assumed that the natural progression of prostate cancer and the sensitivity of a PSA test followed by a biopsy were the same in the United States as in ERSPC–Rotterdam. The predicted prostate cancer incidence peak in the United States was then substantially higher than the observed prostate cancer incidence peak (13.3 vs 8.1 cases per 1000 man-years). However, the actual observed incidence was reasonably reproduced by assuming a substantially lower PSA test sensitivity in the United States than in ERSPC–Rotterdam. For example, for nonpalpable local- or regional-stage cancers (i.e., stage T1M0), the estimates of PSA test sensitivity were 0.26 in the United States versus 0.94 in ERSPC–Rotterdam. We conclude that the efficacy of PSA screening in detecting prostate cancer was lower in the United States than in ERSPC–Rotterdam.

Prostate-specific antigen (PSA) testing was introduced in the United States in 1986 to monitor prostate cancer progression. The test was rapidly adopted for the early detection of prostate cancer, and as a consequence, the incidence of prostate cancer has increased rapidly since 1988, peaking in 1992.[1] The benefits and harms of PSA testing depend on its performance in detecting prostate cancers and on the benefits of consequent early treatment. The performance of PSA testing as a screening test depends on the cutoff level for recommending a biopsy, the compliance to a biopsy recommendation, and the diagnostic accuracy of the biopsies that are performed.

Differences between PSA screening performance with respect to the detection of prostate cancer in a trial and in a population is crucial information for translating the results of a prostate cancer screening trial to a population setting. In this study, we compared PSA screening performance for detecting prostate cancers in the US population with that in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). Because PSA screening performance in the US Prostate, Lung, Colorectal and Ovarian (PLCO) trial may be comparable to PSA screening performance in the US population,[2] the results of this analysis could also provide quantitative explanations for the different mortality results of the ERSPC and PLCO trials.[3-4]

For this analysis, we used the MIcrosimulation SCreening ANalysis (MISCAN) model for prostate cancer,[5-6] which simulates individual life histories and models the development of cancer in individuals as a sequence of tumor states. The model includes 18 detectable preclinical states in the natural history of prostate cancer that are derived from combinations of clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; and T3+, palpable, with extensions beyond the prostatic capsule),[7] differentiation grade (well differentiated, Gleason score 2–6; moderately differentiated, Gleason score 7; and poorly differentiated, Gleason score 8–10),[8] and metastatic stage (local or regional [M0] and distant [M1]).[7] Cancer can progress from each preclinical state to the clinical disease state (i.e., become diagnosed because of symptoms). Preclinical cancers may be detected by PSA screening. Screen-detection depends on the timing of PSA tests and on the test sensitivity. In the MISCAN model, the PSA test and a subsequent biopsy are modeled as a single test; therefore, PSA test sensitivity also depends on whether a positive test is followed by a biopsy. In the model, sensitivity is defined as the probability that a preclinical tumor is detected by a screening test at the time the test is taken. The parameters for PSA test sensitivity are stage-specific because the sensitivity of a test primarily depends on the size of the tumor.

Model parameters, including transition probabilities, mean dwelling times (the time from one preclinical state to another preclinical or clinical state), and stage-specific test sensitivities, are typically estimated as follows. A model is constructed for a specific situation, such as prostate cancer incidence in the United States or in both arms of the ERSPC–Rotterdam. Parameters are then estimated by numerical minimization of the deviance between observed

numbers of cases and the number of cases predicted by the model. Deviances are calculated by assuming Poisson likelihood for incidence data or by assuming a multinomial likelihood for stage distribution data.

In this study, we first developed an ERSPC model that simulated the prostate cancer progression and screening in ERSPC–Rotterdam. Estimates of natural history parameters and test sensitivities were obtained by using the observed detection rates, interval cancer rates, and stage distributions from ERSPC–Rotterdam.[5-6]

Next, to make the model results comparable to observed US data, the population in the model was adjusted to the US population by replacing the birth tables and life tables with US-specific tables, and the screening protocol of ERSPC–Rotterdam was replaced with the frequency of PSA testing in the US population. The frequency of PSA testing in the United States was modeled according to the approach described by Mariotto et al.[9] The frequency of a first PSA test and of repeat tests in the United States, as reproduced in the MISCAN model, is illustrated in Figure 1.4. On average, 80% of the screened men in the United States have a repeat PSA test within 2 years of the previous test.

We considered two US models. In model 1, we investigated the hypothesis that PSA screening in the United States is the same as in ERSPC–Rotterdam. In this model, all prostate cancer–related parameters were the same as in the ERSPC model. In model 2 we investigated the hypothesis that the sensitivity of PSA screening in the United States is lower than that in ERSPC–Rotterdam. In this model, all prostate cancer–related parameters except for the test sensitivity parameters were the same as those in the ERSPC model. US-specific estimates of

Table 3.1: Estimates of sensitivity, detection rate, and deviance for the two US models*

Parameter	Model 1	Model 2
Sensitivity by stage†		
T1M0	0.94	0.26
T2M0	0.94	0.26
T3M0	1	0.27‡
T1M1	0.96	0.84
T2M1	0.97	0.84
T3M1	1	0.84
Detection rate per 1000 screened men		
At first PSA test	62	18
At repeat PSA test	13	12
Deviance	44,727	23,438

*PSA = prostate-specific antigen.

†T1, T2, and T3 are the three clinical T-stages (T1, nonpalpable; T2, palpable, confined to the prostate; and T3, palpable, with extensions beyond the prostatic capsule), M0 is the local or regional stage, and M1 is the distant stage.

‡The range of plausible values is (0.24–0.29). The range of plausible values indicates a range in which the 95% confidence interval will be with near certainty, see Appendix Figure 3.1. Because of restrictions on the sensitivities (sensitivity increases with clinical T-stage and metastatic state), this range can not be calculated for the other parameters

test sensitivities were obtained by using observed age-specific incidence and age-specific stage distribution (local or regional vs distant) in the US population. For estimation of the US-specific parameters, we used data from the Surveillance, Epidemiology, and End Results (SEER) registry for US men aged 50–84 years who were diagnosed with prostate cancer between January 1, 1975, and December 31, 2000. The data were based on the nine core catchment areas (SEER 9) of the SEER registry (http://seer.cancer.gov/). We used the test sensitivity parameter estimates of the ERSPC model as starting values for optimization of the estimates of the US model. The estimated test-sensitivities parameters of the calibrated model are presented in Table 3.1.

In model 1, both the predicted and observed incidence peaks occurred in 1992. However, the predicted prostate cancer incidence peak in the United States was substantially higher than the observed prostate cancer incidence peak (13.3 vs 8.1 cases per 1000 man-years), which suggests a lower detection of prostate cancer in the United States than in ERSPC–Rotterdam (Figure 3.1). In model 2, the predicted incidence peak was the same size as the observed incidence peak (Figure 3.1). However, estimates of test sensitivity were lower in the United States than in ERSPC–Rotterdam. For example, for nonpalpable local- or regional-stage cancers (i.e., stage T1M0), the estimates of PSA test sensitivity were 0.26 in the United States versus 0.94 in ERSPC–Rotterdam (Table 3.1).

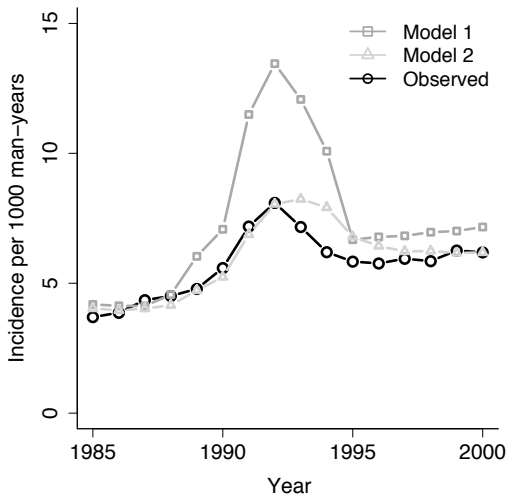


Figure 3.1: Observed (solid) and predicted (dashed) age-adjusted incidence per 1000 man-years for men aged 50–84 years in the US models. In model 1, prostate-specific antigen (PSA) screening in the US population is the same as in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer (ERSPC–Rotterdam). In model 2, the sensitivity of PSA screening is lower in the US population than in ERSPC–Rotterdam.

The lower sensitivity of PSA screening in the United States compared with ERSPC–Rotterdam in model 2 could be due to a higher PSA cutoff level for recommending biopsy in the United States, a lower biopsy compliance rate in the United States, or a lower sensitivity of the biopsies in the United States. The latter possibility is unlikely because more biopsy cores are generally taken in the United States than were taken in ERSPC–Rotterdam. The other two possibilities might explain the lower sensitivity of PSA screening in the United States. A higher PSA cutoff level for recommending biopsy in the United States could follow from the fact that the recommended PSA cutoff level in the United States is 4 ng/mL, whereas the PSA cutoff level in ERSPC–Rotterdam was 3 ng/mL. A lower biopsy compliance rate in the United States could, for instance, indicate that some physicians in the United States might have used a higher PSA cutoff level than recommended (ie, higher than 4 ng/mL) or might have advised a confirmatory PSA test if the first PSA level was elevated. Confirmatory PSA tests would lower the biopsy compliance rate because men with a PSA level higher than the cutoff level at the first test but with a PSA level lower than the cutoff level at the confirmatory test would probably be advised to not have a biopsy; therefore, some men with a PSA level higher than the cutoff level at the first PSA test would not have a biopsy. Pinsky et al.[2] reported a biopsy compliance rate in the PLCO trial of 41% within 1 year of a positive PSA test. They suggested that this biopsy compliance rate is representative of US screening practice given that men with a positive PSA test in the PLCO trial were referred to their personal physician for follow-up. In the screening arm of ERSPC–Rotterdam, biopsies were administered by the screening center at no charge to the subject and reminders for biopsy appointments were sent if necessary, resulting in a biopsy compliance rate of approximately 90%. In model 2, the detection rates at first PSA screening and at repeat PSA screening were 18 and 12 per 1000 screened men, respectively (Table 3.1), which are comparable to the detection rates at the first round of screening (16 per 1000 screened men) and repeat screening (11 per 1000 screened men) in the PLCO trial.[10]

This study has four limitations. First, we did not take into account other factors, such as race, that differed between the US and ERSPC–Rotterdam populations and might influence the detection rates. Approximately 10% of the US population is black, whereas nearly 100% of the ERSPC–Rotterdam population was white. Because the incidence of prostate cancer was higher among black men than among white men during the study period, these racial differences might explain the different detection rates estimated for the two populations. However, the incidence of prostate cancer among whites in the US population was similar to the overall incidence[11] which indicates that the effect of black men on the overall observed incidence was small, as was their effect on the outcomes of this study.

Second, we assumed that the model that we used for the frequency of PSA testing[9] would apply to screening tests. During the construction of that model, all follow-up PSA tests taken after diagnosis as well as PSA tests occurring within three months of a previous PSA test

were eliminated. However, a fraction of the remaining tests might be diagnostic tests that were used to confirm a suspicion for prostate cancer. The size of this fraction is unknown, but including this fraction of diagnostic tests as screening tests would imply that the actual screening rate is lower than in the model.

Third, a weakness of our model is that it fails to explain why prostate cancer incidence in the US dropped so steeply after 1992 (Figure 3.1). In our model, the cancers detected in repeat tests led to a slower decline of incidence after 1992 than what was observed. However, the frequency of repeat PSA testing remained at a level of 30% (Figure 1.4), and it is unclear why these tests detected so little cancer in the US population.

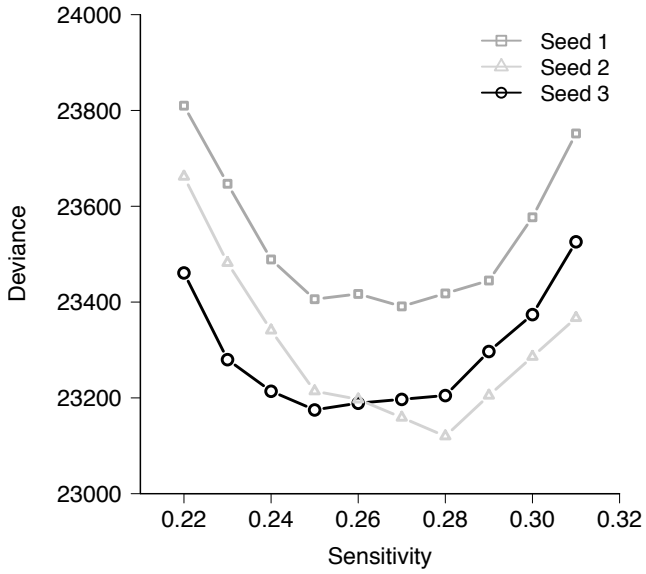
Fourth, we could not compute 95% confidence intervals for the sensitivity parameters: because of random noise in the simulated predictions and restrictions on the sensitivities (sensitivity increases with clinical T-stage and metastatic state), formal 95% confidence intervals are difficult to obtain when using the microsimulation model. However, for fixed values of other model parameters, the range of plausible values for test sensitivity for a local or regional stage tumor in clinical stage T3 (i.e., in state T3M0) was narrow (0.24–0.29). The range of plausible values contains with near certainty a standard computed 95% confidence interval. The calculation of the range of plausible values is presented in Appendix Figure 3.1 in the Appendix.

In conclusion, PSA screening in the United States did not detect as many prostate cancers as PSA screening in ERSPC–Rotterdam because of the lower sensitivity of PSA testing and consecutive biopsy. The consequence of this lower test sensitivity is that the effects of PSA screening in the United States are likely to be different from those observed in the ERSPC–Rotterdam. For example, Draisma et al.[12] noted that the lead time (time by which screening advances diagnosis) and the frequency of overdiagnosis were smaller in the United States than in ERSPC–Rotterdam (mean non-overdiagnosed lead time: 6.9 vs 7.9 years; overdiagnosis frequency: 42% vs 66%), indicating that the harms of PSA testing in the United States, while still substantial, are likely to be less than those in the ERSPC–Rotterdam. The benefits of PSA screening in the United States are also likely to be different from those in ERSPC–Rotterdam. The ERSPC trial has shown that screening for prostate cancer by using PSA tests can reduce prostate cancer mortality;[4] however, we can not directly translate these mortality reductions to the US population because of differences between the two populations, such as the lower sensitivity of PSA screening in the United States. Finally, this analysis also shows quantitatively that it is likely that there is a difference in the sensitivity of the PSA screening (PSA test and consecutive biopsy) in the ERSPC and PLCO trials, which is likely to have contributed to the different outcomes of the trials.

REFERENCES

1. Legler JM, Feuer EJ, Potosky AL, Merrill RM, Kramer BS. The role of prostate-specific antigen (PSA) testing patterns in the recent prostate cancer incidence decline in the United States. *Cancer Causes Control.* Oct 1998;9(5):519-527.
2. 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.* Mar 2005;173(3):746-750; discussion 750-741.
3. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
4. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
5. 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.* Jun 18 2003;95(12):868-878.
6. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
7. Schroder FH, Hermanek P, Denis L, Fair WR, Gospodarowicz MK, Pavone-Macaluso M. The TNM classification of prostate cancer. *The Prostate.* 1992;4:129-138.
8. Gleason DF. Classification of prostatic carcinomas. *Cancer chemotherapy reports.* Mar 1966;50(3):125-128.
9. 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.* May 1 2007;109(9):1877-1886.
10. Grubb RL, 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int.* Dec 2008;102(11):1524-1530.
11. Ries LAG, Harkins D, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2003, National Cancer Institute: http://seer.cancer.gov/csr/1975_2003/. In. Bethesda, MD; 2006.
12. 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.* Mar 18 2009;101(6):374-383.

APPENDIX



Appendix Figure 3.1: Calculating plausible values for the estimate of the parameter for sensitivity for a tumor in clinical stage T3 and in the local or regional stage (M0). Because of the random noise in the simulated predictions and restrictions on the sensitivities (sensitivity increases with clinical T-stage and metastatic state), formal 95% confidence intervals are difficult to obtain. However, for fixed values of other model parameters, the range of plausible values for test sensitivities is narrow (0.24–0.29). The range of plausible values contains with near certainty a standard computed 95% confidence interval. The range of plausible values was calculated as follows. First, we computed deviances for the model for different test sensitivities around the maximum likelihood estimate. In this figure we present the different deviances for the model if we vary the sensitivity for three different random sequences. The deviance was computed as usual as 2 × (the log likelihood of a saturated model minus the log likelihood of the model). In standard maximum likelihood estimation, 95% confidence intervals would be calculated, based on the likelihood ratio test, as the interval between the points where the deviance is 3.84 higher (the 95% quantile of 1 df for a chi-square distribution) than the minimum at the maximum likelihood estimate. As this figure shows, the deviances are approximate quadratic functions that have a minimum between 0.25 and 0.28, but random noise prevents the application of the standard 3.84 rule. However, the graphs also show that the likelihood of values outside the range of 0.24–0.29 must be small, which is the range of plausible values.

Chapter 4

How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening

Elisabeth M. Wever, Gerrit Draisma, Eveline A.M. Heijnsdijk and Harry J. de Koning

Medical Decision Making. 2011;31(4):550-558.
Reprinted with permission from Sage Publications.

ABSTRACT

Background

Simulation models are essential tools for estimating benefits of cancer screening programs. Such models include a screening-effect model which represents how early detection by screening followed by treatment affects disease-specific survival. Two commonly used screening-effect models are the stage-shift model, where mortality benefits are explained by the shift to more favorable stages due to earlier detection, and the cure model, where early detection enhances the chances of cure from disease. The objective of this paper is to describe the commonly used screening-effect models and analyze their predicted mortality benefit in a model for prostate cancer screening.

Methods

The MISCAN simulation model was used to predict the reduction of prostate cancer mortality in the European Randomized Study of Screening for Prostate Cancer (ERSPC) Rotterdam. The screening-effect models were included in the model. For each model the predictions of prostate cancer mortality reduction in ERSPC-Rotterdam were calculated. We compared four screening-effect models, which are versions of the stage-shift model or the cure model.

Results

The stage-shift models predicted, after a follow-up of nine years reductions in prostate cancer mortality varying from 38% to 63% for ERSPC-Rotterdam compared with a 27% observed in the overall ERSPC. The cure models predicted reductions in prostate cancer mortality varying from 21% to 27%.

Conclusions

The differences in predicted mortality reductions show the importance of validating models to observed trial mortality data. Using the stage-shift models to include the effect of screening considerably over-estimated the mortality reduction. Therefore, the stage-shift models should be used with care, especially when modeling the effect of screening for cancers with long lead times, such as prostate cancer.

INTRODUCTION

Screening can be used for the early detection of several types of cancer.[1] For example, mammography is commonly used to detect breast cancer in an earlier stage and Pap smear is used to detect potentially precancerous lesions and prevent cervical cancer.[2] Screening for colorectal cancer with Fecal Occult Blood Test (FOBT), flexible sigmoidoscopy and colonoscopy has been tested in randomized trials.[3] Prostate Specific Antigen (PSA) screening for the early detection of prostate cancer is widespread in the US.[4] Death from cancer might be postponed or prevented if a cancer is found earlier as treatment is received at a less advanced stage. However, this benefit can only be obtained if the treatment that is received is more effective at a less advanced stage.

Models are often used to simulate the progression of cancer and how this progression is affected by screening. Such models are described as natural history models. They have been important tools for estimating benefits and harms of interventions for early detection, for instance the screening of breast cancer,[5-6] cervical cancer[7-8] and colorectal cancer.[9-10] Models have also been used for explaining observed trends in cancer incidence and mortality. For example Etzioni et al.[11] used two models to project that 45% to 70% of the observed decline in prostate cancer mortality by year 2000 could plausibly be attributed to the stage-shift due to screening. Also, Berry et al.[12] showed with seven models that mammography screening and treatment have helped to reduce the rate of death from breast cancer in the United States.

Natural history models include a screening-effect model. A screening-effect model determines how early detection by screening followed by treatment affects the progression of cancer, specifically on how diagnosing and treating patients at a less advanced stage affects prostate cancer mortality. Screening-effect models may be summarized generally into two types: stage-shift models attribute the better prognosis because of early detection by screening to the shift to a less advanced stage with the corresponding more favorable stage-specific survival. Cure models assume that early detection is followed by curative treatment that is either successful in preventing cancer-specific mortality or unsuccessful, i.e. not changing the time and cause of death of patients. Mortality predictions from the models may vary significantly with the type of screening-effect model used. In this paper we study the effect of various stage-shift and cure models on the predicted prostate cancer mortality in the European Randomized Study of Screening for Prostate Cancer (ERSPC) Rotterdam and compare this effect with the results observed in the overall ERSPC.[13]

METHODS

In most simulation models the progression of cancer in individuals is simulated first in the absence of screening (Figure 4.1, Panel B). In the absence of screening, the time of prostate

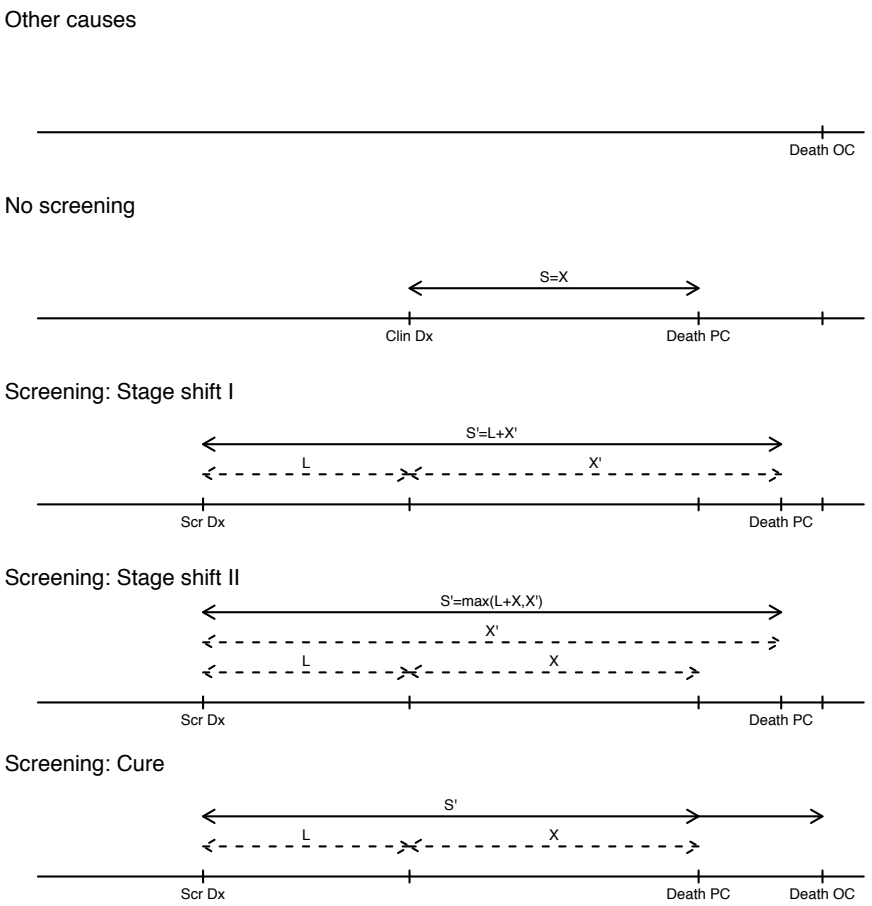


Figure 4.1: Modeling the impact of early detection on disease progression. Each panel shows relevant events in an individual life history on a time line. Panel A indicates the time of death from other causes than prostate cancer (Death OC). Panel B presents the prostate cancer history in the absence screening. Survival *S* after clinical diagnosis (Clin Dx) equals a random variable *X*, drawn from a survival curve specific for stage at diagnosis. Panels C to E illustrate the impact of early detection by screening in the various models. Panel C: With the Stage-shift 1 model, survival *S'* after detection by screening (Scr Dx) equals the sum of lead time *L* and a random variable *X'* drawn from a survival curve specific for the stage at the time of early detection. Panel D: With the Stage-shift 2 model *X'* is generated similarly, but survival after detection by screening *S'* is taken to be the maximum of the original survival (*L*+*X*) and the new *X'*. Panel E: In the cure models, survival *S'* is up to death from other cause (Death OC) with probability *c* (cure) and equals the original survival corrected for lead time (*L*+*X*) with probability 1-*c* (no cure).

cancer death is defined by survival, *S*, after clinical diagnosis which equals a random variable, *X*, drawn from a survival curve specific for stage at diagnosis.

These natural history models assume a preclinical phase that precedes the time of clinical diagnosis during which tumors can be detected by screening. How early detection by screening affects the modeled disease progression depends on the screening-effect model. The screening-effect model is a sub-model that takes into account that persons diagnosed with cancer early by screening may have a better cancer-specific survival compared to persons clinically diagnosed when symptoms appear. In general, persons with cancers found by screening have better survival in part because there is a lead time component in the survival and also because finding and treating cancers earlier might postpone the time of death of cancer. This study considers four different screening-effect models: two stage-shift models and two cure models. Figure 4.1 illustrates the different models in Panel C to E, and Table 4.1 presents the model parameters for the specific models.

Stage-shift models

Stage-shift models attribute the improved prognosis associated with early detection by screening to the shift to a less advanced cancer stage with the corresponding more favorable stage-specific survival. A simple application of this principle is that a survival time from the

Table 4.1: Parameters and data source for each screening-effect model.

Model	Parameters	Parameter value by stage at diagnosis			Source
		Local/Regional Gleason score <=7	Local/Regional Gleason score >7	Distant Any Gleason score	
Stage-shift 1, 2A, 2B	Stage specific survival at*				SEER prostate-cancer-specific survival in pre PSA era (1983-1986) with hazard ratio 0.65 from Bill-Axelson et al.[14]
	5 year	0.98	0.73	No effect (0.45)*	
	10 year	0.90	0.49	No effect (0.29)*	
Cure 1A	Stage specific probability of cure	0.33	0.17	No effect	Ratio of long term prostate-cancer-specific survival of non-treated and treated prostate cancer in pre-PSA era (SEER 1983-1986) with hazard ratio 0.65 from Bill-Axelson et al.[14] (S(20)0.65-S(20))/(1-S(20)0.65)
Cure 1B	Stage specific probability of cure	0.42	0.23	No effect	Calibration to the observed 27% mortality reduction in ERSPC at nine years of follow-up.[13]
Cure 2A	Stage specific probability of cure	0-0.69: increasing with lead time up to 5 years	0-0.37: increasing with lead time up to 5 years	No effect	Calibration to the 27% observed mortality reduction in ERSPC at nine years of follow-up.[13]

*Value in parentheses is the prostate-cancer-specific survival probability for patients detected with prostate cancer by clinical diagnosis. If prostate cancer is detected by screening in the distant stage it is assumed that screening has no effect on the survival.

moment of screen-detection is generated from a stage (at detection) specific survival function, which is the same survival function as for men clinically diagnosed with cancer. This simple application may lead to implausible consequences. Namely, that patients' live may be shorter in the case of early detection by screening than in the case of no screening. This may be plausible, for example, if treatment is associated with mortality. However, because the mortality due to treatment is usually relatively small, this situation does not arise in most models. Most stage-shift models employ an additional assumption to avoid implausible results of reduced life-expectancy with screening. In this study we consider the following two stage-shift models.

In *Stage-shift model 1* no death from cancer is allowed during lead time, i.e. before the time of clinical diagnosis in the absence of screening. Technically, we assume that persons whose cancers are detected by screening have a cancer-specific survival, *X'*, which depends on the stage, age and treatment at screen-detection and that the survival time starts after the lead time, *L*, i.e. the survival starts at the expected time of clinical diagnosis in the absence of screening. This implies that the survival, *S'*, after detection by screening equals the sum of lead time, *L*, and the random variable, *X'*, drawn from a survival curve specific for the stage at the time of early detection (Figure 4.1, Panel C). In this screening-effect model the possibility of patients dying during lead time is ruled out and the possibility of dying before the expected time of dying had patients not been screened, is substantially decreased.

In *Stage-shift model 2,* no death from cancer is allowed before the time of death from cancer in the absence of screening. Specifically, persons whose cancers are detected by screening have a cancer-specific survival, *X'*, which depends on the stage, age and treatment at screen-detection but survival time starts at the time of screen-detection. However, if this new survival would imply dying from cancer before the expected time of death in the absence of screening such persons are assumed to die of cancer at the same time of death as in the absence of screening. This implies that the survival, *S'*, after detection by screening is taken to be the maximum of the new survival, *X'*, and the original survival, *L+X*, (Figure 4.1, Panel D).

We consider two versions of *Stage-shift model 2*. In *Stage-shift model 2A* we assume that the survival at screen-detection and the survival at clinical-detection are independent. In *Stage-shift model 2B* we assume that the two survivals are dependent. Technically, the dependency is included by generating the survival at screen-detection and the survival at clinical-detection with the same random number, i.e. using the same quantile from both survival functions. In *Stage-shift model 2* survival is truncated at the survival time as in the case that the individuals had not been screened, implying that men who would die of prostate cancer in the absence of screening can additionally benefit from the independency.

Cure models

In the cure models a fraction of the tumors detected by screening are cured because the tumors are treated earlier. Patients will not die of cancer if they are cured, but if they are not

cured, their date and cause of death are not changed by early detection. This implies that survival *S'* is up to death from other causes with probability *c* (cure) and equals the original survival corrected for lead time, *L*+*X*, with probability 1-*c* (no cure), (Figure 4.1, Panel E).

In this study we consider two cure models. In *Cure model 1* we assume that the cure rate (the probability of cure) is constant and we consider two estimates for the cure rate. The cure rate is either estimated from the ratio of long term disease specific survival of treated versus untreated cancer (*Cure model 1A*), or from the mortality reductions observed in randomized trials (*Cure model 1B*).

In *Cure model 2* the cure rate depends on the predicted lead time (time by which screening advances diagnosis). For this study we assume that the cure rate increases linearly with the lead time, *L,* for the first 5 years and that the cure rate is constant after a lead time of 5 years. Namely, for lead times < 5 years the cure rate = *L* x *c*/5 and for lead time ≥ 5 years cure rate = *c*, where the parameter *c* is estimated by calibrating the model to mortality predictions observed in a trial. Note however that a non-linear relationship can also be assumed between the cure rate and the lead time.

Comparing the screening-effect models

The MISCAN prostate cancer model was used to simulate the progression of prostate cancer and screening in the ERSPC-Rotterdam.[15] Models were constructed using the different screening-effect models. The prostate cancer mortality reduction in the ERSPC-Rotterdam by follow-up year was estimated using each of these models. The prostate cancer mortality reduction was calculated as the ratio of the number of prostate cancer deaths prevented by screening and the number of prostate cancer deaths in the absence of screening. Predictions of mortality reduction at nine years of follow-up of the ERSPC-Rotterdam were compared to the published mortality reduction of 27% at nine years of follow-up in the overall ERSPC.[13] All results were calculated for the screened men in the core age group, men between age 55 and 69 at randomization, of the ERSPC.

The MISCAN prostate cancer model

The MISCAN prostate cancer model is briefly described here. A more detailed description is available at http://cisnet.cancer.gov/prostate/profiles.html. The MISCAN prostate cancer model is a micro-simulation program which simulates the progression of cancer in individuals as a sequence of tumor states. The individual life histories are first simulated in the absence of screening. Prostate cancer may develop from no prostate cancer to a clinically diagnosed cancer through one or more screen-detectable preclinical stages. In each preclinical stage, a tumor may grow to the next clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; and T3+, palpable, with extensions beyond the prostatic capsule), dedifferentiate to a higher Gleason score (well differentiated: 2–6; moderately differentiated: 7; and poorly differentiated: 8–10;), or give rise to symptoms and become clinically diagnosed. The time spent

in the current stage is generated from a Weibull distribution, where the parameters depend on the current stage. The choice of the next stage is determined by transition probabilities. In addition, there is a risk that a tumor in the local-regional stage (M0) will develop into distant disease (M1). The transition from the local-regional stage to the distant disease is modeled by a stage- and grade-specific hazard function. The model thus includes 18 detectable preclinical states that are derived from combinations of clinical T-stages, differentiation grades, and metastatic stages.

The parameters of the progression of prostate cancer were estimated by constructing models for the ERSPC-Rotterdam and calibrating the model to observed data: baseline incidence (1991) and stage distribution (IKR92/93) in the Netherlands; incidence, Gleason and stage distributions in the control arm; and detection rates, interval cancer rates, Gleason and stage distributions in the screen arm. Note that the parameters of the progression of prostate cancer before diagnosis were estimated independent of survival data. A detailed description of the model's progression component is given by Draisma et al.[15-16]

After men are clinically diagnosed with prostate cancer, treatment is assigned. The men can receive three treatment types: active surveillance, radical prostatectomy and radiation therapy. The same treatment fractions have been assumed in the model as in the ERSPC-Rotterdam, depending on age, Gleason score and clinical T-stage. For the corresponding treatment the time of death from prostate cancer is obtained using prostate-cancer-specific survival curves. Bill-Axelson et al.[14] estimated a relative risk of 0.65 for death from prostate cancer for men treated with radical prostatectomy compared to patients with no initial treatment (active surveillance). According to this result, we assumed that men receiving radical prostatectomy or radiation therapy have a relative risk of 0.65 compared to active surveillance for local-regional cancers and that those men receiving active surveillance experience the baseline prostate cancer survival. For distant prostate cancer it is assumed that treatment has no effect on survival, implying that irrespective of the treatment type, all men diagnosed with prostate cancer in the distant stage have a survival generated from the corresponding baseline prostate-cancer-specific survival curve.

The baseline prostate-cancer-specific survival curves have been estimated on the basis of SEER (Surveillance, Epidemiology and End Results) data in the pre-PSA (prostate-specific antigen) era, specifically of cases diagnosed between 1983 and 1986. The survival curves were modeled using Poisson regression with grade, stage, age and treatment type as explanatory variables. To assign the prostate-cancer-specific survival curves in our model we assumed that Gleason score of 7 or less corresponds to grade well/moderately differentiated and that Gleason score more than 7 corresponds to grade poor/undifferentiated.

Screening-effects are then modeled by superimposing screening on the life histories in the absence of screening. Preclinical cancers may be detected by screening, depending on the frequency and the sensitivity of the screening test for the specific preclinical state. The PSA test and subsequent biopsy are modeled as a single test. In accordance with the

screening in the ERSPC-Rotterdam, it is assumed that the men in the screening arm had their first PSA test between November 1993 and December 1999 and that there were 3 screening rounds with a time-interval of 4 years between the screening rounds. The time of prostate cancer death for the men screen-detected with prostate cancer is determined using each of the screening-effect models described before in turn.

For each screening-effect model, the parameters and the data from which the parameter estimates are obtained are presented in Table 4.1. In the stage-shift models, the effect of early detection by screening is entirely determined by the stage-shift and the corresponding prostate-cancer-specific survival curves.

In the cure models, the effect of early detection by screening depends on the cure rate which has to be estimated. In *Cure model 1A* we estimated the cure rate based on the ratio of 20 years prostate-cancer-specific survival of treated versus untreated cancer (Table 4.1). The cure rate estimates were 0.33 for Gleason scores less or equal to 7 and 0.17 for Gleason scores greater than 7.

In *Cure model 1B* and *2*, the cure rate was estimated assuming in the model a mortality reduction of 27% in the ERSPC-trial Rotterdam after a follow-up of nine years for men who were actually screened. The mortality reduction of 27% was observed in the overall ERSPC-trial. We additionally assumed that the ratio between the cure rates given the Gleason scores is the same as in *Cure model 1A*. For *Cure model 1B* the cure rate estimates were 0.42 for Gleason scores less or equal to 7 and 0.23 for Gleason scores greater than 7. For *Cure model 2* the estimate for the cure parameter, *c*, was 0.69 for Gleason scores less or equal to 7 and 0.37 for Gleason scores greater than 7.

RESULTS

The predicted prostate cancer mortality reductions by follow-up time in the ERSPC-Rotterdam are presented in Figure 4.2. *Stage-shift model 1* predicted a prostate cancer mortality reduction of 45% after a follow-up of 9 years. *Stage-shift model 2* predicted a prostate cancer mortality reduction after a follow-up of 9 years of 63% for the model assuming that the survival at screen-detection and the survival at clinical-detection are independent (A) and 38% for the model assuming that the survival at screen-detection and the survival at clinical-detection are dependent (B). *Cure model 1A*, which assumes a constant cure rate based on the ratio of long term survival of treated versus untreated cancer, gave the lowest mortality reduction of 21% after a follow-up of 9 years. *Cure model 1B*, which assumes a constant cure rate based on the ERSPC outcome, had a predicted prostate cancer mortality reduction of 27% after a follow-up of 9 years. *Cure model 2* also had a predicted prostate cancer mortality reduction of 27% after 9 years, as this model was also calibrated to the 27% prostate cancer mortality reduction observed in the overall ERSPC. However, *Cure model 2*, which assumes a

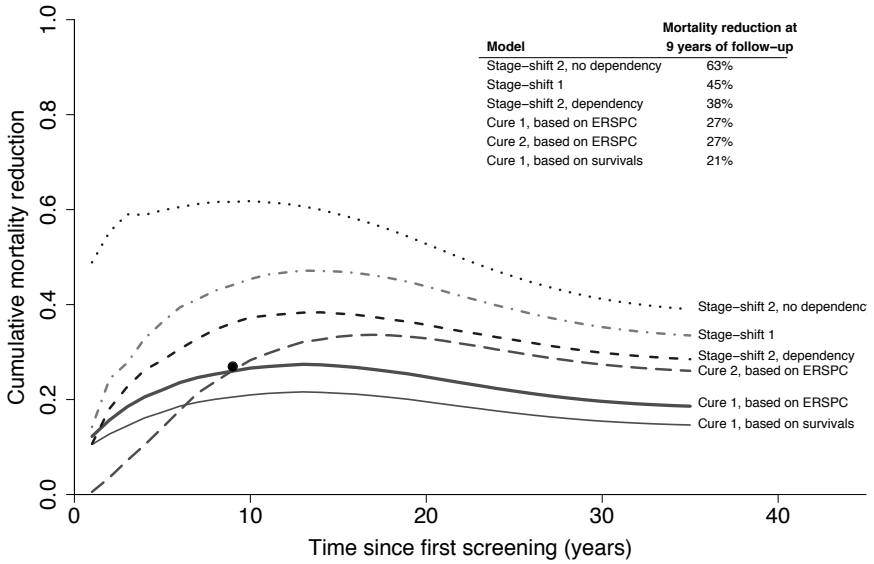


Figure 4.2: Predicted prostate cancer mortality reduction by follow-up time for men aged 55-69 at the first screening, using the different sub-models for the effect of screening. The dot shows the observed 27% mortality reduction at nine years of follow-up in the ERSPC. *Stage-shift model 1*: the cancer-specific survival starts after the lead time. *Stage-shift model 2*: the survival starts at screen-detection but is taken as the maximum of the new and the original survival. *Cure model 1*: assuming a constant cure rate. *Cure model 2*: assuming a cure rate that depends on the lead time.

cure rate that depends on lead time gave lower prostate cancer mortality reductions in the years before the 9th follow-up year but higher prostate cancer mortality reduction in the following year.

Table 4.2 presents the lifetime probabilities of death from prostate cancer by stage at diagnosis for men with screen-detected prostate cancer. The effect of screening is shown by the reduction in the risk of death from prostate cancer when the cancer is detected earlier by screening. The effect of early detection by screening for local-regional cancers with Gleason score less or equal to 7 was larger in the stage-shift models (0.10-0.13) than in the cure models (0.05-0.09). Conversely, the effect of screening for local-regional cancers with Gleason score greater than 7 was larger in general in the cure models (0.08-0.11) than in *Stage-shift model 1 and 2B* (0.03-0.04). Only *Stage-shift model 2A*, in which the survival starts at screen-detection but is taken as the maximum of the new and the original survival and where independence was assumed between the survival at screen-detection and the survival at clinical detection, gave a larger effect (0.17) than the cure models and the other stage-shift models.

No benefit was associated with early detection by screening for cancers found in the distant stage. The relative benefit, the risk of death from prostate cancer prevented by screening compared to the risk from prostate cancer in case of no screening, was limited for cancers found in the local-regional stage with Gleason score greater than 7. However, there was a

considerable relative benefit for cancers found in the local-regional stage with Gleason score less or equal to 7.

Table 4.2: Predicted lifetime risks of prostate cancer mortality in men screen-detected with prostate cancer by stage and model.

		Model					
Stage at detection	Risk	SS 1*	SS 2A†	SS 2B‡	Cure 1A§	Cure 1B‖	Cure 2¶
Local/regional and Gleason score <= 7	Death from prostate cancer without screening	0.14	0.14	0.14	0.14	0.14	0.14
	Death from prostate cancer with screening	0.03	0.01	0.04	0.09	0.08	0.05
	Death from prostate cancer prevented by screening	0.11	0.13	0.10	0.05	0.06	0.09
Local/regional and Gleason score > 7	Death from prostate cancer without screening	0.48	0.48	0.48	0.48	0.48	0.48
	Death from prostate cancer with screening	0.44	0.31	0.45	0.40	0.37	0.38
	Death from prostate cancer prevented by screening	0.04	0.17	0.03	0.08	0.11	0.10
Distant and any Gleason score	Death from prostate cancer without screening	0.72	0.72	0.72	0.72	0.72	0.72
	Death from prostate cancer with screening	0.72	0.72	0.72	0.72	0.72	0.72
	Death from prostate cancer prevented by screening	0.00	0.00	0.00	0.00	0.00	0.00

*SS 1: Stage-shift model 1, the cancer-specific survival starts after the lead time.
†SS 2A: Stage-shift model 2A, the survival starts at screen-detection but is taken as the maximum of the new and the original survival, assuming no dependency between the new and original survivals.
‡SS 2B: Stage-shift model 2B, the survival starts at screen-detection but is taken as the maximum of the new and the original survival, assuming dependency between the new and original survivals.
§Cure 1A: Cure model, assuming a constant cure rate based on survivals.
‖Cure 1B: Cure model, assuming a constant cure rate based on the ERSPC outcome.
¶Cure 2: Cure model, assuming a cure rate that depends on the lead time and where the parameter estimate is based on ERSPC outcome.

DISCUSSION

The mortality predictions for ERSPC-Rotterdam vary considerably between the stage-shift models and the cure models, which employ different assumptions regarding the effect of early detection by screening. The stage-shift models predicted substantially larger prostate cancer mortality reductions than the 27% observed in the ERSPC-trial.[13]

It is likely that the best way of modeling the effect of early detection is to assign survivals that are based on observed survivals specific to screen-detected cancers and clinically diagnosed cancers. In this case, the survival for screen-detected cancers would already include

both the lead time and the survival benefit because of early detection. However, these data are often not available and therefore additional assumptions that define how early detection by screening affects disease-specific survival are often needed.

The simplest version of the stage-shift model is sometimes used, for example, in models of natural history and screening of colorectal cancer.[17-18] In this simple screening-effect model, if cancers are detected earlier by screening, the survival time from the moment of screen-detection is generated from a stage (at detection) specific survival function, which is the same survival function as for men clinically detected with cancer. These survivals may lead to under-estimates when the lead time is incompletely compensated by the shift to early stages and the corresponding better survival. Colorectal cancer has a relatively short mean lead time of around 2.1 years.[19] Therefore, neglecting part of the lead time component in the survival for screen-detected colorectal cancers might not be a significant concern. However, using this simple stage-shift model in our prostate cancer study would suggest that 90% of men with screen-detected cancers would die earlier of prostate cancer than if these cancers were clinically detected. This implausible result is obtained because part of the lead time component is neglected in the survival for screen-detected cancers, while the mean lead time for prostate cancer is relatively long (5.4-6.9 years in the US population and 7.9 years in the ERSPC Rotterdam trial[20]).

By employing additional assumptions, the stage-shift models used in this study avoid the occurrence of deaths from cancer after early detection which occurs earlier than deaths from cancer if the patients were not screened and were clinically-detected at a later time. *Stage-shift models 1 and 2* used here have previously been used for modeling the effect of early detection by screening for breast cancer.[21-22]

In general, the use of *Stage-shift model 1* leads to mortality reductions that are over-estimated. This is because the lead time component is over-compensated. After the lead time, men with screen-detected cancers are assigned survival specific to their stage at screen-detection, even though their screen detection was some years previously. Consequently, for cancers with a large lead time and considering very few disease stages, this screening-effect model can give mortality reductions that are considerably over-estimated. Since prostate cancer has a large lead time[20] and the survival considers only 4 different stages, *Stage-shift model 1* in this study gave prostate cancer mortality reductions that are greatly over-estimated. Breast cancer on the other hand has a relatively short lead time of 0.83-1.4 years,[19, 23] implying that using *Stage-shift model 1* for modeling the effect of early detection of breast cancer might not be a significant concern.

Etzioni et al.[11] used in two models *Stage-shift model 1* to quantify the contribution of PSA screening to the prostate cancer mortality reduction observed in the US population. They concluded that PSA screening explains 45% to 70% of the mortality decline observed in the US population by the year 2000. The results of our study show that using this screening-effect

model leads to over-estimated mortality reductions and that therefore probably less of the mortality decline can be attributed to PSA screening.

Whether using *Stage-shift model 2,* where the survival starts at screen-detection but is taken as the maximum of the new and the original survival, gives over-estimates or under-estimates for mortality reductions is not trivial. The magnitude of the effect of screening depends on the lead time and the survivals. In the extreme cases, cancers with short lead times and long survivals are more likely to have large predicted benefits of screening, whereas cancers with long lead times and short survivals are more likely to have small predicted benefits of screening. Therefore, whether the screening-effect of the model is correct depends jointly on the lead time and the survivals. In this particular study of prostate cancer screening, the *Stage-shift model 2* gave mortality predictions that are over-estimated.

Replicating this study for other cancers might not lead to the conclusion that the stage-shift models over-estimate the effect of screening for those cancers. However, this study demonstrates that the screening-effect models should not be used without comparing the mortality predictions with observed mortality data.

In the absence of screening-trial results, the cure rate could be estimated from the ratio of long term prostate-cancer-specific survival of treated versus untreated cancer (*Cure model 1A*). Using this cure rate estimate gave a predicted prostate cancer mortality reduction that is somewhat lower than the observed reduction in the ERSPC-trial. Estimating the cure rate (*Cure model 1B and 2*) by calibrating the model to the observed prostate cancer mortality reduction in the overall ERSPC, leads to survival benefits that result in a 27% prostate cancer mortality reduction in the ERSPC-Rotterdam. In *Cure model 1* we assumed a constant cure rate and in *Cure model 2* we assumed a cure rate that increases with length of the lead time. It is unclear whether by using one of these two screening-effect models the mortality reduction before and after the ninth year of follow-up is modeled well. It is also unclear whether the effect of screening is modeled correctly by stage at diagnosis (Table 4.2). In this study we assumed an increasing cure rate for the lead time years 1 to 5, however we could have also assumed a different relationship between the cure rate and the lead time. For example, we could have assumed that the cure rate increases slowly for long lead times and faster for shorter lead times. However, since mortality reductions before and after the ninth year of follow-up are not yet available, it is not possible to validate the cure models completely.

In our models, the parameters of the progression of prostate cancer up to clinical diagnosis and screen-detection were estimated first and the cure parameter was estimated subsequently. With sufficient follow-up data of screen-detected and clinically diagnosed patients it would be useful to estimate these parameters jointly by calibrating the model to observed incidence data, detection rates and mortality data.

A limitation of this study is that we predicted prostate cancer mortality reduction in the ERSPC-Rotterdam and compared it with the observed 27% prostate cancer mortality reduction in the overall ERSPC and not with the one observed in the ERSPC-Rotterdam. However,

we do not expect the reduction in prostate cancer mortality in the ERSPC-Rotterdam to be very different, as according to the results of the ERSPC-trial the rate ratio for death from prostate cancer changes very little when the Rotterdam section is excluded from the ERSPC.[13]

In conclusion, different screening-effect models can be used to model the effect of screening and these different models gave different outcomes in this study. The first message of this study is that it is important to make explicit what assumptions are made about the effect of early detection by screening on survival. The second message is that trial data are essential for the validation of the screening-effect models. The advantage of the cure model is that the cure parameter could be calibrated to the observed 27% prostate cancer mortality reduction in the ERSPC-trial. Using the stage-shift models to include the effect of screening considerably over-estimated the mortality reduction. Therefore, the third message is that stage-shift models should be used with care when modeling the effect of screening of cancers with long lead times, such as prostate cancer.

REFERENCES

1. Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW. Cancer screening in the United States, 2010: a review of current American Cancer Society guidelines and issues in cancer screening. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):99-119.
2. Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R. Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *J Natl Cancer Inst.* Nov 21 2001;93(22):1704-1713.
3. Benson VS, Patnick J, Davies AK, Nadel MR, Smith RA, Atkin WS. Colorectal cancer screening: a comparison of 35 initiatives in 17 countries. *Int J Cancer.* Mar 15 2008;122(6):1357-1367.
4. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *J Natl Cancer Inst.* Oct 7 2009;101(19):1325-1329.
5. de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. *Eur J Cancer.* Jan 2009;45(1):127-138.
6. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine.* Nov 17 2009;151(10):738-747.
7. Mandelblatt JS, Lawrence WF, Womack SM, et al. Benefits and costs of using HPV testing to screen for cervical cancer. *JAMA.* May 8 2002;287(18):2372-2381.
8. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *J Natl Cancer Inst.* Feb 6 2002;94(3):193-204.
9. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Boer R, Wilschut J, Habbema JD. At what costs will screening with CT colonography be competitive? A cost-effectiveness approach. *Int J Cancer.* Mar 1 2009;124(5):1161-1168.
10. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine.* Nov 4 2008;149(9):659-669.
11. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control.* Mar 2008;19(2):175-181.
12. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine.* Oct 27 2005;353(17):1784-1792.
13. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
14. 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.* Aug 20 2008;100(16):1144-1154.
15. 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.* Jun 18 2003;95(12):868-878.
16. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
17. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM. Cost-effectiveness of screening for colorectal cancer in the general population. *JAMA.* Oct 18 2000;284(15):1954-1961.
18. Lansdorp-Vogelaar I, van Ballegooijen M, Zauber AG, Habbema JD, Kuipers EJ. Effect of rising chemotherapy costs on the cost savings of colorectal cancer screening. *J Natl Cancer Inst.* Oct 21 2009;101(20):1412-1422.
19. Wu D, Rosner GL, Broemeling LD. Bayesian inference for the lead time in periodic cancer screening. *Biometrics.* Sep 2007;63(3):873-880.
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.* Mar 18 2009;101(6):374-383.
21. Mandelblatt J, Schechter CB, Lawrence W, Yi B, Cullen J. 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. *Journal of the National Cancer Institute.* 2006(36):47-55.
22. Plevritis SK, Sigal BM, Salzman P, Rosenberg J, Glynn P. A stochastic simulation model of U.S. breast cancer mortality trends from 1975 to 2000. *Journal of the National Cancer Institute.* 2006(36):86-95.
23. Chen Y, Brock G, Wu D. Estimating key parameters in periodic breast cancer screening-application to the Canadian National Breast Screening Study data. *Cancer epidemiology.* Aug 2010;34(4):429-433.

Chapter 5

The impact of PLCO control arm contamination on perceived PSA screening efficacy

Roman Gulati, Alex Tsodikov, Elisabeth M. Wever, Angela B. Mariotto, Eveline A.M. Heijnsdijk, Jeffrey Katcher, Harry J. de Koning and Ruth Etzioni

Cancer Causes & Control. 2012;23(6):827-835.
Reprinted with permission from Springer Science and Business Media.

ABSTRACT

Purpose

To quantify the extent to which a clinically significant prostate cancer mortality reduction due screening could have been masked by control arm screening (contamination) in the Prostate, Lung, Colorectal, and Ovarian (PLCO) trial.

Methods

We used three independently developed models of prostate cancer natural history to conduct a virtual PLCO trial. Simulated participants underwent pre-trial screening based on population patterns. The intervention arm followed observed compliance during the trial then resumed population screening. A contaminated control arm followed observed contamination during the trial then resumed population screening, while an uncontaminated control arm discontinued screening upon entry. We assumed a clinically significant screening benefit, applied population treatments and survival patterns, and calculated mortality rate ratios relative to the contaminated and uncontaminated control arms.

Results

The virtual trial reproduced observed incidence, including stage and grade distributions, and control arm mortality after 10 years of complete follow-up. Under the assumed screening benefit, the three models found that contamination increased the mortality rate ratio from 0.68–0.77 to 0.86–0.91, increased the chance of excess mortality in the intervention arm from 0–4% to 15–28%, and decreased the power of the trial to detect a mortality difference from 40–70% to 9–25%.

Conclusions

Our computer simulation models indicate that contamination substantially limited the ability of the PLCO to identify a clinically significant screening benefit. While the trial shows annual screening doesn't reduce mortality relative to population screening, contamination prevents concluding whether screening reduces mortality relative to no screening.

INTRODUCTION

The survival benefit due to prostate-specific antigen (PSA) screening for prostate cancer remains uncertain and controversial. Long-awaited results from the European Randomized study of Screening for Prostate Cancer (ERSPC) and the US-based Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial have not resolved this critical issue. A significant reduction in prostate cancer mortality was observed in the ERSPC, which reported a prostate cancer mortality rate ratio for the intervention arm versus the control arm of 0.80 after a median of 9 years of follow-up.[1] This result was not replicated in the PLCO study, which reported corresponding mortality rate ratios 1.15 after 10 years (complete for 92% of participants) and 1.09 after 13 years (complete for 57% of participants).[2]

Reconciling these apparently conflicting results is challenging because they reflect the performance of different screening protocols in different populations under different circumstances. A key difference is that the ERSPC was conducted in a population that was relatively screening-naïve whereas the PLCO trial was conducted in a population in which screening was already becoming established. Not only did at least 45% of PLCO trial participants enter the trial with a history of screening,[2] but those randomized to the control arm continued to be screened during the course of the trial at least as often as in the general population, averaging 2.7 routine PSA tests over the trial's 6-year intervention period.[3] Thus the control arm was heavily contaminated by screening, and rather than comparing screening versus no screening, the PLCO trial effectively evaluated organized annual screening versus population screening.[2, 4] A recent evidence review for the US Preventive Services Task Force noted the potential role of contamination in explaining the PLCO's non-significant prostate cancer mortality rate ratio and the trend towards excess mortality in the intervention arm.[5] However, existing statistical methods to quantify the magnitude of its effect,[6] as have been applied to the ERSPC,[7] require stratifying outcomes within contaminated and uncontaminated subgroups, stratifications that may be sensitive to the definition of contamination and may not be possible using the PLCO surveys upon which published contamination estimates were based.[3]

In this article we use three independently developed mathematical and computer simulation models to represent a virtual PLCO trial with both contaminated and uncontaminated control arms. The models were developed to study prostate cancer progression, screening, detection, and survival in the US population[8-10] as part of the Cancer Intervention and Surveillance Modeling Network (CISNET);[11] here we adapt them to study PLCO trial results. We assume a clinically significant benefit of screening and quantify the extent to which contamination masks the benefit in terms of its effect on the mortality rate ratio, on the chance of observing excess mortality in the intervention arm, and on the power of the trial to detect the assumed benefit.

METHODS

Overview of CISNET prostate models

In this article, we combine previously developed CISNET models of prostate cancer natural history with an updated model of population screening patterns and published data on the PLCO trial design. Each model is based on a specification of natural history that captures disease onset, progression, detection in the absence of screening, and survival. The natural history concepts differ across models, but all models were calibrated using common datasets on (a) population demographics: men aged 50–84 in the core 9 registries of the Surveillance, Epidemiology, and End Results (SEER) program[12] and (b) PSA screening patterns: a reconstruction[13] using Medicare claims and responses from the National Health Interview Survey (NHIS) in 2000 to estimate the probability of a first PSA test by birth cohort and calendar year, which we extended using responses from the NHIS in 2005. The calibrated models reproduce age-adjusted SEER incidence trends before and during the PSA era by SEER stages (local-regional or distant) and Gleason grade categories (2–7 or 8–10). Detailed model descriptions are available on the CISNET website;[11] here we briefly review the key model features and primary data sources for calibration.

The FHCRC model is PSA-based; we estimate PSA growth for cases and non-cases using serial screening data from the control arm of the Prostate Cancer Prevention Trial,[14] and we represent the risk of disease onset as proportional to age and the risk of progression to metastasis and to disease detection in the absence of screening as proportional to PSA. Because the FHCRC model explicitly generates PSA trajectories, the PSA value at any testing event is known, and the consequences of a positive test, namely referral to biopsy, receipt of biopsy, and biopsy result given disease status, are generated as events in the individual's disease history. The likelihood of compliance with biopsy referral is based on studies conducted within the PLCO trial population,[15-16] and the sensitivity of biopsy to detect occult disease is assumed to increase over time in accordance with the diffusion of extended biopsy protocols during the PSA era.[17] Progression risks given PSA are then estimated via maximum likelihood so that the model matches incidence trends in the SEER population.[9]

The MISCAN model is state-based; disease progresses through a sequence of states defined by stage and grade. The stage and grade transition rates are estimated using data from the Rotterdam section of the ERSPC;[8] the progression to disease detection in the absence of screening and the sensitivity of the PSA test to detect disease within each state are based on recalibration of these model parameters to reproduce disease incidence trends in the SEER population.[18]

The UMICH model of natural history consists of three states: healthy, preclinical, and clinical (diagnosis in the absence of screening). This simple model is overlaid with a model of the time from onset to detection (delay time), which depends on the distribution of age and calendar year of diagnosis, and a model of stage and grade at diagnosis, which depends

on the delay time.[10] Putting together these models of age/year incidence trends and stage/grade distributions with inputs on SEER incidence and US screening patterns produces a composite model for incidence in the SEER population.[19]

Adapting the CISNET models to conduct a virtual PLCO trial

From 1993 to 2001, the PLCO trial enrolled 76,693 men aged 55–74 at 10 US study centers and randomized them to either an intervention or control arm.[20] Men in the intervention arm were offered annual PSA testing for 6 years and digital rectal exam testing for 4 years. This design provided 90% power to detect a 20% reduction in prostate cancer mortality.[21] Under the design assumptions about noncompliance and contamination, this corresponds to a 27% mortality reduction,[21] which was deemed to be clinically significant. In order to achieve this level of power, trial investigators expected over 300 deaths in each arm after 10 years of complete follow-up. In practice, after 10 years of follow-up (complete for 92% of participants), the incidence of prostate cancer death was very low, with 98 and 85 deaths in the intervention and control arms and a mortality rate ratio of 1.15 (95% confidence interval 0.86 to 1.55).

We adapt the CISNET models to conduct a virtual PLCO trial by simulating a cohort of men with observed age distributions[20] in each arm at randomization. The models generate disease natural histories in the absence of screening, assign birth-cohort-specific ages at PSA screenings based on our extended model of the general US population,[13] and determine all-cause mortality from US life tables; the generation algorithm ensures all men are alive and undiagnosed at entry into the trial. Lacking information on observed recruitment and for simplicity, we assume all enrollment into the virtual trial occurs on January 1, 1993.

Men randomized to the intervention arm switch to annual PSA screening with random 85% compliance at each scheduled screen[20] for the trial intervention period then resume population screening. Men randomized to the control arm switch to a 20% higher intensity of screening than the general population during the trial to match previous estimates of PLCO control arm contamination (including reproducing the average 2.7 routine PSA tests during the trial[3]) then resume population screening. Men in a third uncontaminated control arm discontinue PSA screening upon entry into the trial. Screening patterns in the three arms are illustrated in Figure 5.1. Digital rectal exam testing is not explicitly modeled since it is already reflected in the way the models project diagnosis in the absence of PSA screening.

All models assume 35% of control arm men who were tested and had PSA > 4 ng/mL, like their intervention arm counterparts,[15] received a biopsy. All models assume that the trial's sextant biopsy had 80% sensitivity to detect preclinical tumors;[22-23] for simplicity, all models assume this constant 80% sensitivity during and after the trial for both the intervention and control arms. All models randomly assign 84% of men diagnosed with local-regional stage disease to curative treatment based on proportions of SEER local-regional stage cases receiving radical prostatectomy (SEER codes 50, 58, 60, and 68 prior to 1997 and 50 and 60 begin-

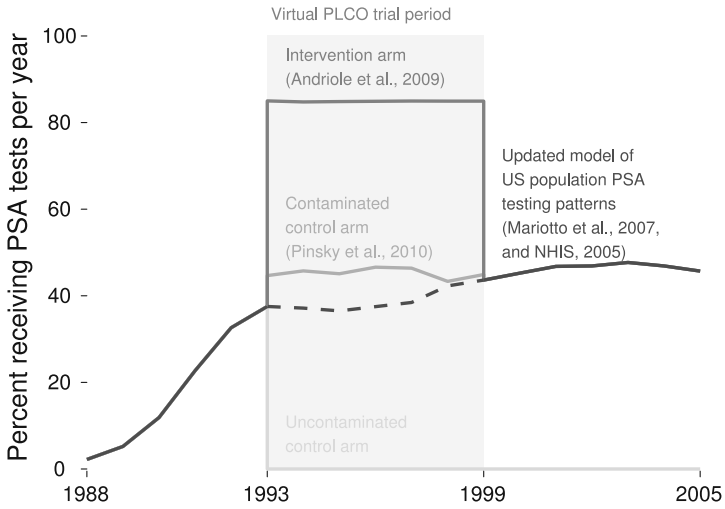


Figure 5.1: Percent of men in each arm of the virtual PLCO trial receiving PSA tests per year. All simulated participants undergo pre-trial screening (solid black line before virtual trial period) based on our model of PSA testing patterns in the general US population applied to the virtual PLCO cohort. The intervention arm follows annual screening subject to observed compliance for the virtual trial period (dark gray line) then resumes population screening (solid black line after virtual trial period). The contaminated control arm follows a higher intensity of screening (solid medium gray line) than in the general population (dashed black line) for the virtual trial period then resumes population screening. The uncontaminated control arm discontinues screening upon entry into the trial (solid slight gray line).

ning in 1998) or radiation therapy (SEER categories beam radiation, radioactive implants, radioisotopes, combination of beam with implants or isotopes, or radiation with method or source unspecified) for trial cohort ages and years. All models generate prostate cancer survival from clinical diagnosis in the absence of screening or treatment benefits using a common Poisson regression model fit to SEER data for cases diagnosed in 1983–1986, just prior to the advent of PSA screening. This baseline prostate survival is improved for non-metastatic cases who receive curative treatment using a hazard ratio of 0.62 based on the Scandinavian trial of radical prostatectomy versus watchful waiting.[24]

To reflect the benefit of screening, all models project results under two different mechanisms that approximate a clinically significant 27% mortality reduction. The stage-shift mechanism implies that cases shifted to an earlier stage and/or grade by screening earn the prostate cancer survival associated with the earlier tumor characteristics, with a corresponding increase in life-expectancy. When a tumor is detected early, but with the same tumor characteristics as in the absence of screening, there can still be a modest survival advantage due to detection at an earlier age. The cure rate mechanism assumes a common constant cure rate among screen-detected non-metastatic cases who would have died of their disease in the absence of screening. Once screen detected, these cases are cured and go on to die of other causes in the presence of screening.[25] A preliminary study using the MISCAN model

found that a constant cure rate of 35% allowed that model to reproduce the ERSPC's non-attendance-adjusted 27% mortality reduction[1] based on data from the Rotterdam center, and all models use this value for the cure rate. Under either mechanism, overdiagnosed cases (i.e., cases detected by screening who would not otherwise have been diagnosed during their lifetimes) do not receive any benefit since they die of other causes both in the presence and absence of screening.

Each model simulates the virtual trial 100 times. To confirm that the model reasonably replicates the trial, we compare model-generated incidence, stage and grade distributions, and control arm mortality with published results. Following validation, we use the models to project mortality and calculate (a) mortality rate ratios of the intervention arm relative to the contaminated and uncontaminated control arms after 10 and 13 years of complete follow-up and corresponding 95% confidence intervals assuming a Poisson distribution for the number of deaths in each arm, (b) the percent of trials with excess mortality in the intervention arm, and (c) the percent of trials in which the 95% confidence interval for the mortality rate ratio excludes 1, i.e., the power of the trial.

RESULTS

Figure 5.2 presents observed and model-projected cumulative incidence in the intervention and control arms for the three models. The models closely reproduced observed disease incidence, with projected incidence in the uncontaminated control arm considerably lower than in the contaminated control arm.

After 10 years of follow-up (complete for 92% of participants), PLCO investigators reported 4% of intervention arm and 5% of control arm cases were clinical stage III–IV; the models estimated 3–5% and 5–7% in these groups and 10–17% in the uncontaminated control arm were projected to be diagnosed in distant stage. Similarly, trial investigators reported 9% of intervention arm and 12% of control arm cases had Gleason score 8–10; corresponding model estimates were 10–20% and 14–20% in these groups and 19–32% in the uncontaminated control arm.

Despite closely reproducing incidence patterns and stage and grade distributions, the models substantially overprojected prostate cancer mortality. Figure 5.3 shows observed and model-projected cumulative mortality in the intervention and control arms for the three models. Prior studies of the PLCO population have indicated that not only was this population healthier than the general population,[26] but it also had a higher level of education, higher socio-economic standing, and a higher proportion of whites. These differences could have predisposed the PLCO population towards lower prostate cancer mortality than that in the general prostate cancer case population. In addition, the common Poisson model of baseline prostate cancer survival was fit using SEER cases diagnosed in 1983–1986 before the

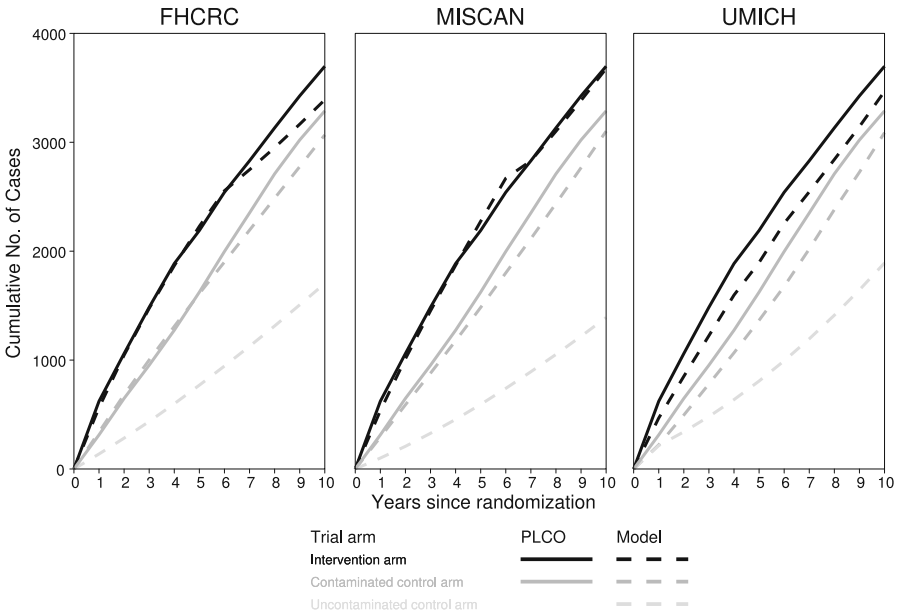


Figure 5.2: Observed and model-projected cumulative prostate cancer incidence in each arm of the virtual PLCO trial. The figure illustrates cumulative number of prostate cancer diagnoses in the intervention arm (black), contaminated control arm (dark gray), and uncontaminated control arm (light gray) observed in the trial (solid lines) and projected by the models (dashed lines).

PSA test was introduced. By the start of the trial, PSA had become an established approach for disease surveillance after diagnosis leading to considerably earlier secondary treatment. For this reason, and possibly also due to other improvements in disease management, it is possible that by the start of the trial prostate cancer survival had improved even before applying screening and primary treatment benefits.

To account for lower-than-expected prostate cancer mortality, all models introduced a hazard ratio to improve baseline prostate cancer survival. Each model estimated a hazard ratio so that the model-generated mortality after applying screening and treatment benefits matched observed mortality in the control arm. The estimated hazard ratios were 0.39 (FHCRC), 0.50 (MISCAN), and 0.37 (UMICH). All prostate cancer mortality results presented below reflect model-generated outcomes using the modified prostate cancer survival models.

Figure 5.4 plots a sorted random sample of mortality rate ratios of the intervention arm relative to contaminated and uncontaminated control arms versus the observed mortality rate ratio after 10 years of complete follow-up. We note that both observed and model-projected 95% confidence intervals are relatively wide, indicating high variability in both actual and virtual trials. This is likely due to the rarity of the event of prostate cancer death despite the large sample size in each arm.

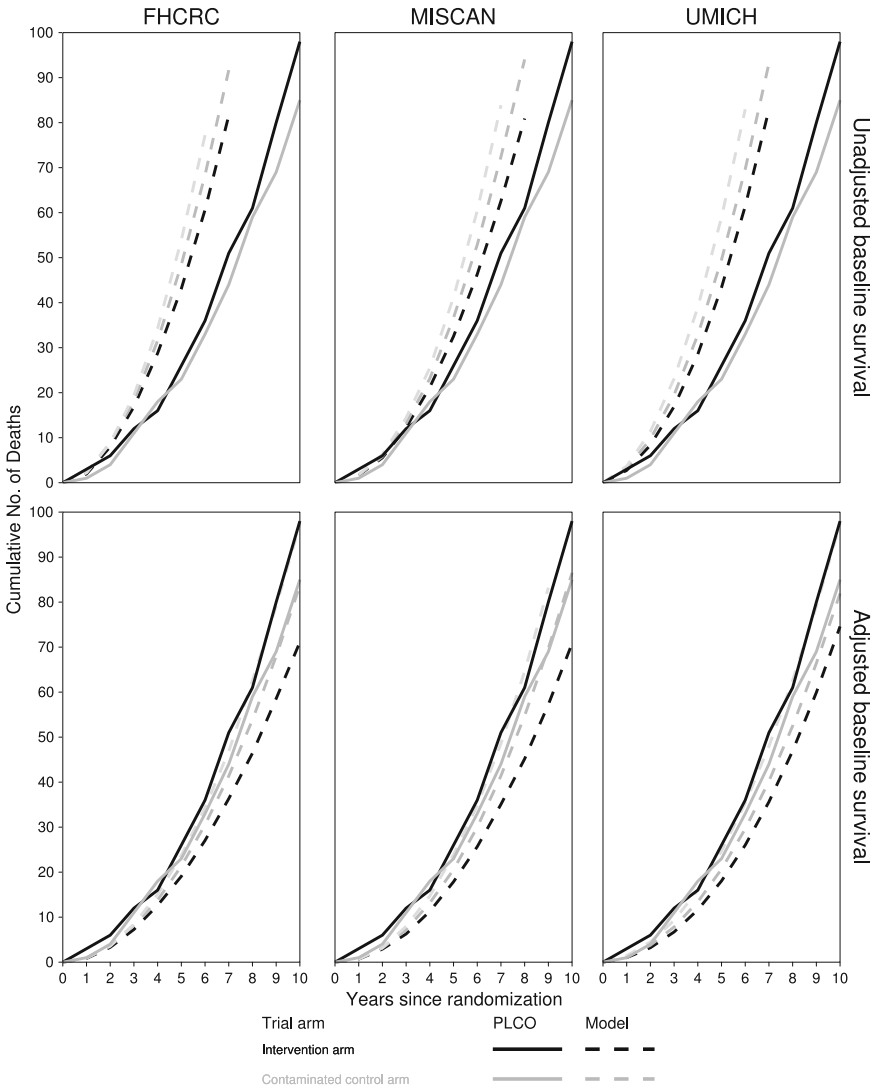


Figure 5.3: Observed and model-projected cumulative prostate cancer mortality in each arm of the virtual PLCO trial. The figure illustrates cumulative number of prostate cancer deaths in the intervention arm (black), contaminated control arm (dark gray), and uncontaminated control arm (light gray) observed in the trial (solid lines) and projected by the models (dashed lines). Model projections assume a stage-shift benefit for screening and a protective benefit of curative treatment. Results are presented separately for unadjusted (top panels) and adjusted (bottom panels) baseline prostate cancer survival to account for lower-than-expected observed mortality.

Table 5.1 summarizes the average projected mortality rate ratio over 100 virtual trials, the percent of trials with excess mortality in the intervention arm, and the power of the trial. After 10 years of complete follow-up, the average mortality rate ratios relative to the uncontaminated control arm across the three models were 0.68–0.77 under the stage-shift and

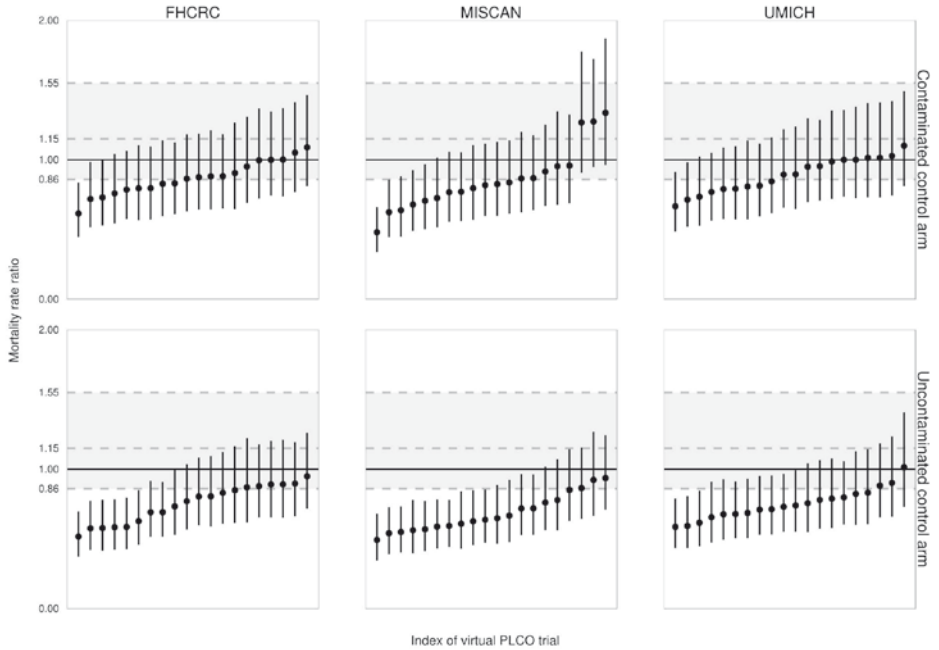


Figure 5.4: Observed and model-projected mortality rate ratios relative to contaminated and uncontaminated control arms. The figure illustrates ratios of mortality rates of the intervention arm relative to the contaminated and uncontaminated control arms and corresponding 95% confidence intervals observed in the trial (dashed gray horizontal lines) and projected by the models (solid black vertical lines) after 10 years of complete follow-up. The model projections represent a sorted random sample of 20 mortality rate ratios and 95% confidence intervals are based on assumed Poisson distributions for the number of deaths.

Table 5.1: Model-projected average mortality rate ratio of the intervention arm relative to the contaminated and uncontaminated control arms over 100 virtual trials, percent of trials with excess mortality in the intervention arm, and percent power of the trial. Results are reported separately under assumed stage-shift and cure rate mechanisms of screening benefit after 10 and 13 years of complete follow-up.

Screening benefit	Years of follow-up	Control arm	Mortality rate ratio			Excess mortality in intervention arm			Power of the trial		
			FHCRC	MISCAN	UMICH	FHCRC	MISCAN	UMICH	FHCRC	MISCAN	UMICH
Stage-shift	10	Uncontaminated	0.75	0.68	0.77	0	0	4	40	70	41
		Contaminated	0.87	0.86	0.91	15	15	28	15	25	9
	13	Uncontaminated	0.75	0.66	0.79	0	0	2	60	90	51
		Contaminated	0.87	0.85	0.93	15	15	29	15	25	9
Cure rate	10	Uncontaminated	0.71	0.80	0.71	0	10	2	60	45	58
		Contaminated	0.81	0.95	0.88	10	30	21	25	5	13
	13	Uncontaminated	0.73	0.81	0.71	0	5	0	80	58	79
		Contaminated	0.85	0.97	0.89	5	42	18	20	11	15

0.71–0.80 under the cure rate. These values are consistent with a clinically significant 27% mortality reduction. Contamination increased the average mortality rate ratios to 0.86–0.91 under the stage-shift and to 0.81–0.95 under the cure rate. Contamination also increased the chance of excess mortality in the intervention arm from 0–4% to 15–28% under the stage-shift and from 0–10% to 10–30% under the cure rate. And contamination decreased the power of the trial to detect the assumed screening benefit from 40–70% to 9–25% under the stage shift and from 45–60% to 5–25% under the cure rate. Results are broadly consistent across models, across the two assumed mechanisms of screening benefit, and across 10 or 13 years of complete follow-up. In an additional sensitivity exercise (results not shown), we found that applying population screening to the contaminated control arm without an adjustment to reflect higher intensity of screening than the general population during the trial yielded very similar impacts of contamination.

DISCUSSION

The perceived negative finding from the PLCO trial that PSA screening does not save lives has generated a storm of questions, particularly in light of the positive finding from the ERSPC. Some studies comparing the two trials have argued that screening in the control arm of the PLCO trial could have been sufficient to mask a clinically significant benefit or to produce a false negative result.[27-30] The impact of contamination on the trial results has not previously been quantified but is critical to proper interpretation of the trial and synthesis of its results with those of the ERSPC and other studies.

In this article we use previously developed models of prostate cancer natural history and, assuming early detection confers a clinically significant survival benefit, quantified the impact of control arm contamination on relative mortality rates in a virtual PLCO trial. We find that control arm contamination substantially equalized mortality rates in the two arms, creating a nontrivial probability of excess mortality in the intervention arm and lowering the power of the trial to detect a mortality difference between the two arms. These impacts remain substantial after accounting for the lower-than-expected mortality in each arm.

Our results depend somewhat on the design assumptions when implementing the virtual trial. We assumed trial enrollment occurs on January 1, 1993, rather than being spread out between 1993 and 2001. A consequence is that we may underproject pre-trial screening; however, the three models estimated that 39–41% of virtual trial participants had at least 1 PSA test in the 3 years before entry, which is similar to the 45% estimated by trial investigators.[2] We ignored changes in trial protocol that increased the number of tests from 4 to 5 in 1994 and from 5 to 6 in 1995 and that excluded men with more than 1 PSA in the preceding 3 years starting in 1995. Conversely, our updated model of PSA testing in the US population, intensified to capture increased screening during the trial, implied that 68% of control arm

participants received at least 1 routine PSA during the trial, which is only slightly lower than the 74% estimated by PLCO investigators.[3] Thus the virtual PLCO involved more screening in the intervention arm and less screening in the control arm than occurred in the actual trial, suggesting that our projected impact of contamination may be conservative. We assumed that intervention and control arm participants resumed population screening after the trial intervention period. We assumed SEER primary treatment patterns for detected cases. And we improved baseline prostate cancer survival to account for lower-than-expected prostate cancer mortality in the control arm. Because these assumptions apply equally to intervention and control arms, they likely have little effect on our projections of relative mortality rates.

Our results also depend on the assumed screening benefit. We chose a level of benefit deemed to be clinically significant by trial investigators[21] and consistent with that observed in the ERSPC. A smaller benefit would diminish while a greater benefit would increase the impact of contamination. Under this assumed benchmark level of benefit, the projected impact of contamination is robust to the assumed mechanism by which early detection confers benefit, to whether evaluation occurs after 10 or 13 years of complete follow-up, and to the three models of underlying disease natural history.

A more complete reconciliation of ERSPC and PLCO results is underway. In collaboration with investigators from both trials, we are using the CISNET models to quantify impacts of different background incidence, pre-trial screening, screening frequencies, PSA thresholds for test positivity, biopsy frequencies, and treatment distributions. While quantifying the contributions of these factors will provide deeper insights into reconciling the trial results, the role of PLCO contamination promises to remain a critical factor.

In summary, the PLCO trial results support the conclusion that organized annual screening does not reduce mortality relative to the level of screening in the control arm. However, because the level of screening in the control arm was substantial, the trial results do not support concluding that screening does not reduce mortality relative to no screening. Instead, a clinically significant benefit could have been masked by contamination, preventing simple synthesis with ERSPC results and limiting the value of the trial for informing recommendations about whether men should undergo PSA screening. Nonetheless, the PLCO trial remains an invaluable resource for learning about prostate cancer biology and epidemiology, PSA distributions and growth patterns, characteristics of the PSA test, and comparative effectiveness of screening strategies with varying testing frequencies.

REFERENCES

1. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
2. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst.* Jan 18 2012;104(2):125-132.
3. Pinsky PF, Blacka A, Kramer BS, Miller A, Prorok PC, Berg C. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial. *Clin Trials.* Aug 2010;7(4):303-311.
4. Berg CD. The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial: the prostate cancer screening results in context. *Acta oncologica (Stockholm, Sweden).* Jun 2011;50 Suppl 1:12-17.
5. Chou R, Croswell JM, Dana T, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine.* Dec 6 2011;155(11):762-771.
6. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med.* May 15 1997;16(9):1017-1029.
7. Roobol MJ, Kerkhof M, Schroder FH, et al. Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *European Urology.* Oct 2009;56(4):584-591.
8. 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.* Jun 18 2003;95(12):868-878.
9. 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.* Oct 2010;11(4):707-719.
10. Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Stat Med.* Aug 30 2006;25(16):2846-2866.
11. National Cancer Institute, Cancer Intervention an Surveillance Modeling Network. http://cisnet.cancer.gov/.
12. Surveillance, Epidemiology and End Results (SEER). http://seer.cancer.gov/.
13. 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.* May 1 2007;109(9):1877-1886.
14. Goodman PJ, Thompson IM, Jr., Tangen CM, Crowley JJ, Ford LG, Coltman CA, Jr. The prostate cancer prevention trial: design, biases and interpretation of study results. *J Urol.* Jun 2006;175(6):2234-2242.
15. Grubb RL, 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int.* Dec 2008;102(11):1524-1530.
16. Pinsky PF, Crawford ED, Kramer BS, et al. Repeat prostate biopsy in the prostate, lung, colorectal and ovarian cancer screening trial. *BJU Int.* Apr 2007;99(4):775-779.
17. Eichler K, Hempel S, Wilby J, Myers L, Bachmann LM, Kleijnen J. Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. *J Urol.* May 2006;175(5):1605-1612.
18. 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.* Mar 3 2010;102(5):352-355.
19. Chefo S, Tsodikov A. Stage-specific cancer incidence: an artificially mixed multinomial logit model. *Stat Med.* Jul 10 2009;28(15):2054-2076.
20. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
21. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials.* Dec 2000;21(6 Suppl):273S-309S.
22. Babaian RJ, Toi A, Kamoi K, et al. A comparative analysis of sextant and an extended 11-core multisite directed biopsy strategy. *J Urol.* Jan 2000;163(1):152-157.
23. Presti JC, Jr., Chang JJ, Bhargava V, Shinohara K. The optimal systematic prostate biopsy scheme should include 8 rather than 6 biopsies: results of a prospective clinical trial. *J Urol.* Jan 2000;163(1):163-166; discussion 166-167.
24. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.

25. Wever EM, Draisma G, Heijnsdijk EA, de Koning HJ. How does early detection by screening affect disease progression?: modeling estimated benefits in prostate cancer screening. *Med Decis Making.* Jul-Aug 2011;31(4):550-558.
26. Pinsky PF, Miller A, Kramer BS, et al. Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Epidemiol.* Apr 15 2007;165(8):874-881.
27. Cetin K, Beebe-Dimmer JL, Fryzek JP, Markus R, Carducci MA. Recent time trends in the epidemiology of stage IV prostate cancer in the United States: analysis of data from the Surveillance, Epidemiology, and End Results Program. *Urology.* Jun 2010;75(6):1396-1404.
28. Croswell JM, Kramer BS, Crawford ED. Screening for prostate cancer with PSA testing: current status and future directions. *Oncology (Williston Park, N.Y.* May 2011;25(6):452-460, 463.
29. Schroder FH, Roobol MJ. ERSPC and PLCO prostate cancer screening studies: what are the differences? *European Urology.* Jul 2010;58(1):46-52.
30. Studer UE, Collette L. What can be concluded from the ERSPC and PLCO trial data? *Urologic oncology.* Nov-Dec 2010;28(6):668-669.

Part 2

Predicting clinical outcomes of prostate cancer progression, screening and treatment

Chapter 6

What if I don't treat my PSA-detected prostate cancer? Answers from three natural history models

Roman Gulati, Elisabeth M. Wever, Alex Tsodikov, David F. Penson, Lurdes Y.T. Inoue, Jeffrey Katcher, Shih-Yuan Lee, Eveline A.M. Heijnsdijk, Gerrit Draisma, Harry J. de Koning and Ruth Etzioni

Cancer Epidemiology, Biomarkers & Prevention. 2010;20(5):740-50.
Reprinted with permission from American Association for Cancer Research.

ABSTRACT

Background

Making an informed decision about treating a prostate cancer detected following a routine prostate-specific antigen (PSA) test requires knowledge about disease natural history, such as the chances that it would have been clinically diagnosed in the absence of screening and that it would metastasize or lead to death in the absence of treatment.

Methods

We use three independently developed models of prostate cancer natural history to project risks of clinical progression events and disease-specific deaths for PSA-detected cases assuming they receive no primary treatment.

Results

The three models project that 20–33% of men have preclinical onset; of these 38–50% would be clinically diagnosed and 12–25% would die of the disease in the absence of screening and primary treatment. The risk that men under age 60 at PSA detection with Gleason score 2–7 would have been clinically diagnosed in the absence of screening is 67–93% and would die of the disease in the absence of primary treatment is 23–34%. For Gleason score 8–10 these risks are 90–96% and 63–83%.

Conclusions

Risks of disease progression among untreated PSA-detected cases can be nontrivial, particularly for younger men and men with high Gleason scores. Model projections can be useful for informing decisions about treatment.

Impact

This is the first study to project population-based natural history summaries in the absence of screening or primary treatment and risks of clinical progression events following PSA detection in the absence of primary treatment.

INTRODUCTION

Choosing the optimal management strategy for a newly diagnosed localized prostate cancer is based, at least in part, upon our understanding of the natural history of the disease in the absence of any aggressive treatment intervention. Useful information about disease natural history for men clinically diagnosed and initially untreated before the widespread adoption of the prostate-specific antigen (PSA) test is available from population-based cohort studies,[1-2] yet relatively little information is available for men diagnosed following a routine PSA test.[3] While the prognosis for screen-detected tumors appears to be better than that for clinically diagnosed tumors,[4] predicting the natural history of a specific PSA-detected tumor is complicated by overdiagnosis and by the lead time associated with the test.

A large proportion of PSA-detected cancers are overdiagnosed, i.e., they would never have progressed to a symptomatic state or been clinically diagnosed in the absence of the test. By definition, an overdiagnosed tumor has an entirely different prognosis than a non-overdiagnosed one. The clinical challenge is to determine whether a given case is overdiagnosed at the time of diagnosis. Further complicating the situation, even if we are able to identify a non-overdiagnosed cancer, prognosis depends critically on the lead time, which is the time by which diagnosis is advanced by screening. Lead times can be highly variable across patients primarily due to heterogeneity of the disease.

Unfortunately, there is no sure way to assess whether a tumor is overdiagnosed or to predict its lead time in clinical practice. Consequently, once a cancer has been detected by screening, it is typically treated, altering its natural history. We can then no longer observe whether or when it would have progressed in the absence of treatment. This data limitation has spawned the development of model-based approaches for inferring lead time, overdiagnosis, and future survival from observed data on disease-specific incidence and deaths. For example, Nicholson and Harland[5] and Parker et al.[6] projected prostate cancer survival for PSA-detected cases using epidemiologic models based on published incidence and survival data. While these studies provided important insights about prognosis, they relied on simplifying assumptions about the populations under study, screening practices, and/or lead time distributions.

An alternative modeling approach is provided by Etzioni et al.,[7] Draisma et al.,[8] and Tsodikov et al.,[9] who developed more biological models of prostate cancer natural history. These models consist of a series of transitions between a healthy state and the clinico-pathologic stages of disease. While these transitions are generally unobservable, the models quantitatively link the respective transition probabilities with resulting observable rates of stage- and grade-specific disease incidence. Calibration of the models to observed disease incidence data allows for estimation of these transition probabilities. Superimposing screening on the calibrated models allows for projection of overdiagnosis frequencies, probabilities of disease progression, and lead time distributions.

With the goals of making transparent all assumptions underlying these more complicated models, strengthening the robustness of modeling methods, and coordinating common input datasets, these three groups joined together to form the prostate working group of the Cancer Intervention and Surveillance Modeling Network (CISNET). In addition to facilitating deeper insights about disease natural history, the comparative modeling approach provides a degree of robustness to model specification since each modeling group makes different assumptions about the mechanisms of disease progression.

The CISNET prostate working group previously collaborated to quantify the contribution of screening to the population declines in prostate cancer mortality[10] and to reconcile differing estimates of overdiagnosis rates and mean lead times.[11] In this paper, we briefly review these models and use them to examine lifetime risks of, mean ages at, and mean years between key disease progression events in the absence of screening and primary treatment. We then project risks of clinical progression events for PSA-detected localized prostate cases who do not receive curative treatment. In addition, we project 20-year prostate cancer and non-prostate cancer survival by age and Gleason score for these cases. Few studies have published detailed information about prostate cancer natural history in the PSA era absent screening and primary treatment. And, to our knowledge, no study has systematically projected population-based risks of clinical progression outcomes following PSA detection or uncertainty in associated survival were the disease to be left untreated.

METHODS

To facilitate comparison, each natural history model was calibrated to the same incidence data: men aged 50–84 in years 1975–2000 in the core 9 registries of the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program.[12] Missing stage and grade information was imputed assuming that it was missing completely at random. To address the upward drift in Gleason scoring of well to moderately differentiated disease during these years,[13] final projections were reported using Gleason score categories 2–7 and 8–10. To disentangle incidence of clinical and screen diagnoses, each model relied on a common retrospective reconstruction of PSA screening patterns in the US population.[14] Also, each model used US life tables to generate non-prostate cancer survival and a common Poisson regression model for prostate cancer survival (adapted from[1]) fit to data from men clinically diagnosed in 1983–1986, just prior to the widespread dissemination of PSA screening. Each model guarantees individual survival during his lead time, and actual dates and causes of death are assigned based on the earlier of disease-specific and other-cause survival times. By explicitly standardizing these common elements, differences across models are due entirely to differences in the conceptual mechanisms used to represent the development and progression of the disease.

Table 6.1: Comparison of high-level features across models.

Model feature	FHCRC	MISCAN	UMICH
Implementation	simulation	simulation	analytic
Disease states	(2 stages) × (2 grades) local-regional, distant stage low-moderate, high grade	(3 × 2 stages) × (3 grades) T0–T3 local, distant stage low, moderate, high grade	(2 stages) × (2 grades) local-regional, distant stage low-moderate, high grade
Progression depends on	current PSA level	current disease state	delay time and mode of detection
Stage progression	yes	yes	yes
Grade progression	no	yes	yes
PSA test sensitivity*	output of model	endogenous parameter	endogenous parameter
Biopsy compliance†	estimated from PLCO	combined with PSA sensitivity	combined with PSA sensitivity
Biopsy sensitivity‡	based on literature review	combined with PSA sensitivity	combined with PSA sensitivity

*Pr(PSA positive| Disease)
†Pr(Biopsy recieved | PSA positive, Disease)
‡Pr (Biopsy positive | Biopsy recieved, Disease)

Table 6.1 presents an at-a-glance comparison of the high-level natural history model features and implementation details of PSA screening and biopsy practices. In all models, the development of a new prostate cancer represents preclinical onset, i.e., the first point at which a properly directed needle biopsy would detect the cancer. And in all models the risk of onset depends on age. Given onset, the three models represent disease progression using different numbers of states and different assumptions about how disease can progress. In all models disease can progress from an organ-confined early stage to a distant metastatic stage. In all models disease can be clinically diagnosed, i.e., diagnosed in the absence of PSA screening. Clinical diagnosis could result from a digital rectal exam or from clinical manifestations of advanced disease, such as obstructive voiding symptoms. In the presence of PSA screening, disease can also be detected following a routine PSA test. At any screen event, the MISCAN and UMICH models estimate the probability of a positive biopsy given disease, while the FHCRC model decomposes this probability into three factors:

$$\begin{aligned} Pr(\text{Biopsy positive} \mid \text{Disease}) = {} & Pr(\text{Biopsy positive} \mid \text{Biopsy received, Disease}) \\ & \times Pr(\text{Biopsy received} \mid \text{PSA positive, Disease}) \\ & \times Pr(\text{PSA positive} \mid \text{Disease}) \end{aligned}$$

and combines external data for the first two factors with results from modeled PSA levels for the third factor.

The three natural history models are briefly reviewed below. Detailed descriptions of individual models and a joint report comparing the models are available on the CISNET website.[15]

The FHCRC model

The FHCRC model assumes that a man's PSA level (on a logarithmic scale) rises linearly with age and that it rises faster (i.e., it has a higher slope) beginning at onset of a biopsy-detectable

preclinical tumor. In addition, disease grade is fixed at onset and post-onset PSA rises faster for Gleason score 8–10 than for Gleason score 2–7. The risk of disease onset is formalized as a hazard function that is proportional to age, while risks of transitioning from localized to metastatic states and from latent to symptomatic states are given by hazard functions that are proportional to the current PSA level[16-19]. This dependence of disease progression on PSA levels implies that individuals with faster PSA growth will tend to have shorter intervals until the disease spreads beyond the prostate or becomes clinically diagnosed.

Grade-specific PSA slopes and variances in the FHCRC model were estimated using data from the control arm of the Prostate Cancer Prevention Trial, which conducted annual screening of 18,882 men for up to 7 years.[20] PSA growth parameters were estimated by fitting random effects models to screened cases who had at least four tests. Given the estimated PSA growth parameters, we then estimated the disease transition hazards. To do this, we simulated natural histories and population disease trends under screening and identified the transition hazards that produced modeled disease incidence trends that best matched observed incidence trends by age, year, stage, and grade. Men with PSA levels of 4.0 ng/mL or greater at screening were assumed to receive a prostate biopsy based on age- and PSA-specific biopsy compliance rates observed in the Prostate, Lung, Colon, and Ovarian (PLCO) cancer screening trial, a randomized clinical trial involving PSA screening for 38,350 men.[21] Biopsy sensitivity to detect occult tumors was allowed to improve over calendar years to reflect the dissemination of more extensive biopsy schemes in the late 1990s.[16-17] Given individual PSA trajectories, screening schedules, and biopsy compliance and sensitivity rates in the population, the hazard rate parameters were estimated using a simulated maximum likelihood algorithm to match model-projected incidence with SEER incidence.

Given the estimated PSA growth and hazards for disease transitions, the FHCRC model simulates complete disease histories for a population of individuals in the absence of screening and primary treatment. Natural history summaries of interest are then estimated empirically from this population. To project risks of clinical progression events for screen-detected cases, the model simulates another population of disease histories in the presence of screening but in the absence of curative treatment.

The MISCAN model

The MISCAN prostate cancer model also simulates individual life histories. The development of cancer in individuals is modeled as a sequence of tumor states, where prostate cancer develops from no prostate cancer through one or more screen-detectable preclinical states to a clinically diagnosed cancer. In each localized preclinical state, a tumor may grow to the next clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; and T3+, palpable, with extensions beyond the prostatic capsule), de-differentiate to a higher SEER histologic grade (well differentiated, Gleason score 2–6; moderately differentiated, Gleason score 7; and poorly differentiated, Gleason score 8–10), or give rise to symptoms and become

clinically diagnosed. For these transitions, the time spent in the current state is generated from a Weibull distribution, where the parameters depend on the current state, and the choice of the next state is determined by transition probabilities. Additionally, there is a risk that a tumor in a SEER local-regional stage will develop into SEER distant stage disease. The transition to distant stage is modeled with a T-stage- and grade-specific hazard function. Consequently, the model includes 18 detectable preclinical states in the natural history that are derived from combinations of clinical T-stages, SEER histologic grades, and metastatic stages. The parameters for the natural history model were estimated using data from the Rotterdam section of the European Randomized study of Screening for Prostate Cancer.[8, 22]

The MISCAN model represents the PSA test and subsequent biopsy as a single test with stage-specific sensitivities estimated from observed incidence data. For calibration to the US situation, we re-estimated these sensitivity parameters and estimated an additional stage-specific risk of clinical diagnosis to capture different pre-PSA disease diagnosis patterns in the US as compared with Europe. US-specific estimates for the parameters were obtained by calibrating the model to the observed age-specific incidence and age-specific SEER stage distribution (i.e., local-regional versus distant stage) using maximum likelihood.[23]

As in the FHCRC model, natural history summary measures in the absence of screening and/or primary treatment are calculated empirically from simulated life histories in the absence of these interventions. And as in the FHCRC model, a new population of life histories is generated in the presence of screening to project risks of clinical progression events for PSA-detected cancers in the absence of primary treatment.

The UMICH model

The UMICH model of disease natural history consists of a sequence of analytical models rather than computer simulation algorithms. These models quantify the likelihood of observed disease incidence trends while averaging over distributions of unobserved factors influencing these trends.

The first component models disease incidence by age and calendar year. Disease incidence depends on time of disease onset, the time from onset to clinical diagnosis in the absence of screening (sojourn time), screening schedules, and test sensitivity. The model assumes distributions for the age at onset and the sojourn time with unknown parameters to be fit using population incidence data. As in the MISCAN model, test sensitivity reflects both the diagnostic properties of the test itself and the frequency and sensitivity of any subsequent biopsy. Sensitivity is modeled as increasing with time from disease onset. Screening schedules are random but are based on the reconstructed distribution of PSA screening patterns in the population.[14] The unknown parameters are estimated by averaging over these distributions and calibrating the resulting marginal incidence against observed incidence.[9]

The second component explicitly models disease grade (Gleason score 2–7 and Gleason score 8–10) and stage (SEER local-regional and distant) at diagnosis. These clinical character-

istics depend on disease natural history through the time from disease onset to detection (delay time) and the mode of detection (screen or clinical). Given the distribution of age at diagnosis and calendar year of diagnosis output from the marginal incidence model, a Bayesian argument is used to derive a distribution for the delay time and the mode of diagnosis. The model for stage and grade at diagnosis is a multinomial logistic model, where delay time and mode of diagnosis are covariates.[24] Putting this model together with the marginal incidence model produces a model for age-, stage-, and grade-specific incidence; calibrating this model to observed incidence allows for estimation of the parameters of the model for stage and grade at diagnosis.

An extension of the second component allows stage progression from screen detection to future (counterfactual) clinical detection. First, the fitted age-, stage-, and grade-specific model is run assuming zero test sensitivity to produce counterfactual "data" on the likely age-, stage-, and grade-specific incidence of disease in the absence of screening. Next, these "data" are used to estimate five unknown parameters representing allowable transition probabilities between stages and grades. Thus, under nonzero test sensitivity, the model is a joint model of age, stage, and grade at two points of diagnosis (real screen and counterfactual clinical). Model components are fit by maximum likelihood.

Given the estimated model components, we rely on analytic derivations to directly project probabilities of natural history summaries in the absence of screening and treatment. To project risks of clinical progression events for PSA-detected cases, we generate lead times from the fitted lead time distribution, assign stage and grade at clinical diagnosis conditional on lead time and age, stage, and grade at PSA detection, then generate prostate cancer survival from the common Poisson regression based on cases diagnosed in the pre-PSA era.

Model projections

We first use the three models to estimate lifetime risks of, mean ages at, and mean years between important prostate cancer natural history events in the absence of screening and primary treatment as follows.

Lifetime risks are calculated as proportions of individuals in the model populations with preclinical onset, clinical diagnosis, metastasis before clinical diagnosis, or prostate cancer death prior to non-prostate cancer death. To calculate risks conditional on onset, we restrict to individuals with onset in their lifetimes and calculate the proportion of these individuals that had the event in their lifetimes.

Mean ages at onset, clinical diagnosis, metastasis before clinical diagnosis, and prostate cancer death are projected in the presence of non-prostate cancer death. We average ages at each event among individuals that survive to the event in our model populations.

Mean years from onset to clinical diagnosis, to clinical diagnosis of metastatic disease, and to prostate cancer death are obtained by averaging time from onset to each endpoint among individuals in the model populations that reached that endpoint in their lifetimes. For

example, we calculate mean years from onset to clinical diagnosis by summing these durations among men clinically diagnosed in their lifetimes and divide this total by the number of such men.

For reference, we also provide summary measures for PSA or clinical diagnosis in the presence of observed PSA screening patterns.[14]

Next we project risks of clinical progression events for men who are PSA-detected in SEER local-regional stage in the model populations in the year 2000 (the most recent year to which the models are calibrated). Among these men, we estimate the proportion who would have gone on to be clinically diagnosed in the absence of screening. Note that by definition this proportion is one minus the fraction overdiagnosed. We also estimate the proportion who would have progressed to a metastatic stage prior to clinical diagnosis in the absence of screening and to prostate cancer death in the absence of any immediate or delayed primary treatment. Projections are tabulated by age and grade at PSA detection for all models and by age, grade, and PSA for the FHCRC model. At present, projections by PSA are only available from this model since it explicitly connects PSA growth and disease progression.

Finally, to provide a more complete picture of the risk of prostate cancer mortality if a PSA-detected tumor is left untreated, we project 20-year prostate cancer and non-prostate cancer survival by age and grade at PSA detection. The results reflect prostate cancer survival in the presence of other causes as in Nicholson and Harland[5], Albertsen et al.,[1] and Parker et al.[6]

RESULTS

Figure 6.1 presents SEER observed and model-projected age-adjusted incidence per 100,000 men aged 50–84. Only the UMICH model captures the increasing trend in the pre-PSA years, and all models project a more sudden rise coincident with early PSA screening than was observed in SEER. Nonetheless, all models reproduce the scale of incidence in the pre-PSA period, the peak following the introduction of PSA screening in the late 1980s, and the re-stabilization at a higher level in the late 1990s.

Table 6.2 presents natural history measures in the absence of screening or primary treatments projected by the three models. In general, the models are broadly consistent in the picture they present of prostate cancer natural history. The disease is widespread, progressing to a biopsy-detectable tumor in 20–33% of men. In the absence of early detection or primary treatments, 38–50% of these tumors go on to be clinically diagnosed, 5–9% metastasize by the time of clinical diagnosis, and 12–25% die of the disease. The preclinical period averages 7–14 years, permitting PSA testing to detect the disease often well in advance of clinical diagnosis (on average 4–9 years for men PSA-detected in the year 2000 as shown in the last row of Table 6.2).

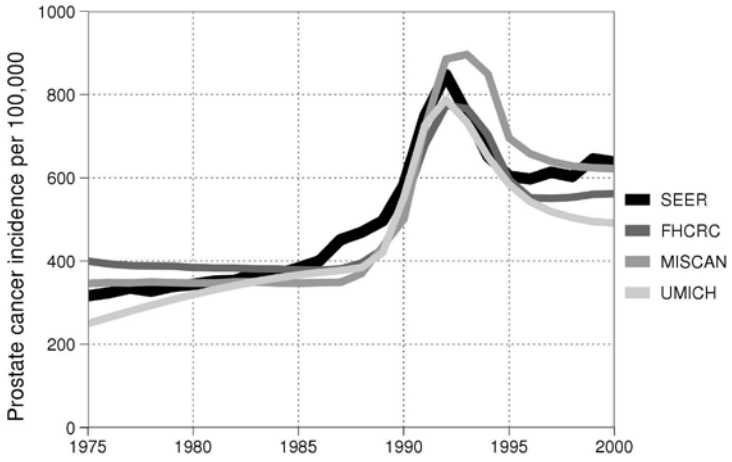


Figure 6.1: Age-adjusted prostate cancer incidence per 100,000 men aged 50–84 observed in the core 9 registries of the Surveillance, Epidemiology, and End Results program of the National Cancer Institute and corresponding projections by the three natural history models.

Table 6.3 presents risks of clinical progression events for PSA-detected cases who receive no primary treatment projected by the three models by age and grade. All results are based on modeled cases screen-detected in SEER local-regional stage in the year 2000. Note that projected risks of all clinical progression events decrease with age; this is to be expected since older men are more likely to die of other causes before the clinical progression event can occur. Also, projected risks tend to be higher when disease is more aggressive, indicated by higher disease grade. In general, even for Gleason score 2–7, all models estimate that at least two thirds of men under age 60 at PSA diagnosis would have gone on to be clinically diagnosed in the absence of screening. The risk that an untreated PSA-detected tumor would have metastasized before clinical diagnosis is 5–23% for men under 60 with Gleason score 2–7 and 22–35% for counterparts with Gleason score 8–10. The risk that an untreated PSA-detected tumor would lead to death in the absence of any primary treatment is 23–34% for men under 60 with Gleason score 2–7 and 63–83% for counterparts with Gleason score 8–10.

Table 6.4 presents corresponding risks by age, grade, and PSA projected by the FHCRC model. We observe that the additional information provided by PSA at a screen detection is important, particularly for older men. For example, the risk that a cancer with Gleason score 2–7 would have gone on to be clinically diagnosed for men aged 65–69 increases from 58% for PSA 4–7 ng/mL at diagnosis to 80% for PSA 10 ng/mL at diagnosis. For the oldest age group, the risk that a cancer with Gleason score 2–7 would have gone on to be clinically diagnosed in the absence of screening, would have metastasized before clinical diagnosis in the absence of screening, or would have become terminal in the absence of primary treatment when PSA is 4–7 ng/mL at diagnosis is 37%, 43%, or 64% the risk faced when PSA is over

Table 6.2: Natural history summary measures projected by the three models.

Measure	FHCRC	MISCAN	UMICH
Lifetime risk of onset*	33	27	20
Lifetime risk of clinical diagnosis	13	12	10
Lifetime risk of metastasis by clinical diagnosis	2	2	1
Lifetime risk of prostate cancer death	4	4	5
Lifetime risk of clinical diagnosis given onset	38	43	50
Lifetime risk of metastasis by clinical diagnosis given onset	7	9	5
Lifetime risk of prostate cancer death given onset	12	14	25
Mean age at onset	65	71	71
Mean age at clinical diagnosis	75	76	81
Mean age at metastasis by clinical diagnosis	74	78	81
Mean age at prostate cancer death	78	80	80
Mean years from onset to clinical diagnosis	14	9	7
Mean years from onset to metastasis by clinical diagnosis	13	13	4
Mean years from onset to prostate cancer death	18	15	8
Lifetime risk of PSA or clinical diagnosis†	14	20	15
Lifetime risk of PSA or clinical diagnosis given onset†	41	73	80
Mean age at PSA or clinical diagnosis†	74	74	75
Mean years from PSA to clinical diagnosis†	6	9	4

*Onset represents the initial development of a biopsy-detectable preclinical tumor.
†These measures are in the presence of observed PSA screening and are included for reference.

Table 6.3: Projected frequencies (%) of clinical progression events for PSA-detected local-regional stage cases who receive no primary treatment by age and grade. Results from three models.

Age group	Gleason grade	Clinical diagnosis			Metastasis by clinical diagnosis			Prostate cancer death		
		FHCRC	MISCAN	UMICH	FHCRC	MISCAN	UMICH	FHCRC	MISCAN	UMICH
50–54	2–7	92	71	93	12	23	5	25	29	34
55–59	2–7	85	67	90	11	19	5	23	26	33
60–64	2–7	77	58	86	11	17	5	20	22	31
65–69	2–7	68	49	80	9	14	5	16	17	28
70–74	2–7	56	40	74	8	10	4	12	12	23
75–79	2–7	44	32	66	6	8	4	8	10	18
80–84	2–7	36	24	59	5	6	3	6	6	13
50–54	8–10	96	96	93	35	23	26	77	83	68
55–59	8–10	90	95	90	33	22	25	69	71	63
60–64	8–10	84	92	86	30	19	24	60	66	56
65–69	8–10	76	89	80	27	21	23	49	55	49
70–74	8–10	66	81	74	23	20	21	39	45	41
75–79	8–10	53	72	66	20	19	19	29	33	32
80–84	8–10	46	64	59	16	17	16	21	23	24

Table 6.4: Projected frequencies (%) of clinical progression events for PSA-detected local-regional stage cases who receive no primary treatment by age, grade, and PSA at detection. Results from the FHCRC model.

Age group	Gleason grade	Clinical diagnosis			Metastasis by clinical diagnosis			Prostate cancer death		
		4–7 ng/mL	7–10 ng/mL	> 10 ng/mL	4–7 ng/mL	7–10 ng/mL	> 10 ng/mL	4–7 ng/mL	7–10 ng/mL	> 10 n g/mL
50–54	2–7	91	92	96	12	13	14	25	25	26
55–59	2–7	82	87	92	11	12	13	21	25	25
60–64	2–7	71	80	88	10	11	12	18	21	24
65–69	2–7	58	69	80	8	9	11	13	16	20
70–74	2–7	44	56	73	6	8	10	9	11	16
75–79	2–7	31	43	64	4	6	10	6	8	12
80–84	2–7	20	31	54	3	4	7	3	4	9
50–54	8–10	92	96	98	33	34	35	72	74	80
55–59	8–10	84	89	95	29	33	35	62	67	73
60–64	8–10	74	81	91	25	28	34	50	55	67
65–69	8–10	62	70	85	21	24	31	35	43	58
70–74	8–10	46	57	77	15	21	27	27	32	47
75–79	8–10	33	43	66	13	14	25	16	24	37
80–84	8–10	21	30	59	8	10	21	9	15	26

10 ng/mL at diagnosis. Corresponding results for a cancer with Gleason score 8–10 are 36%, 38%, and 35%.

Complementing the projected risks that a PSA-detected cancer will lead to death if left untreated, Figure 6.2 presents projected 20-year survival for the same cohort of patients (men aged 50–84 screen detected in SEER local-regional stage in 2000 assuming no primary treatment) by age and grade at PSA detection. Ranges between the minimum and maximum values projected by the three models are represented in darker shaded areas for either cause of death; these regions can be interpreted as uncertainty due to model specification. Agreement across the three models is shown in lighter shaded areas for either cause of death. Among men under age 60 at PSA detection with Gleason score 2–7 disease, the three models project that 4–9% and 15–26% would die of their disease by 10 and 20 years after PSA detection in the absence of treatment. Corresponding projections for men with Gleason score 8–10 disease are 29–43% and 56–68%. Note that similar risks of prostate cancer and non-prostate cancer death and relatively greater inter-model uncertainty create overlapping projections for younger men with Gleason score 8–10.

DISCUSSION

In this article we use three independently developed models to project key events in the natural history of prostate cancer in the absence of screening and primary treatment, namely onset of preclinical biopsy-detectable disease, clinical diagnosis, transition to metastatic

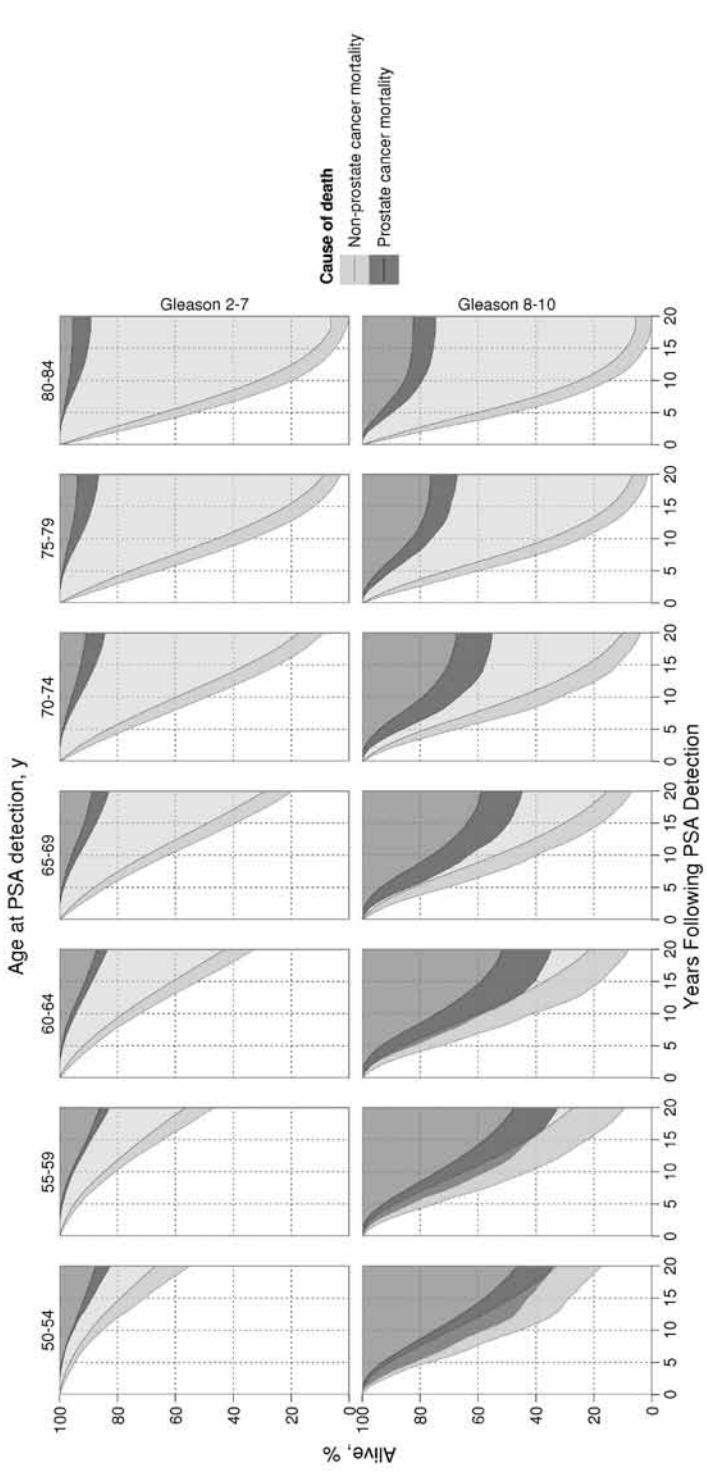


Figure 6.2: Prostate cancer and non-prostate cancer mortality following PSA detection for cases receiving no primary treatment projected by the three models by age and grade at PSA detection. For either cause of death, lighter areas reflect agreement by all three models and darker areas reflect inter-model uncertainty. Similar risks for either cause of death and relatively greater uncertainty about survival create substantial overlap for younger men with Gleason 8–10.

disease by the time of clinical diagnosis, and disease-specific mortality. The models also project frequencies of clinical progression events following PSA detection of local-regional stage disease in the year 2000. All models are estimated using common inputs for observed PSA screening patterns and are calibrated to reproduce US incidence trends among men aged 50–84 in the years 1975–2000. By using population incidence data to inform about underlying disease natural history, we provide unique insights regarding the distributions of unobservable events in disease progression.

Despite each model achieving projections that are reasonably consistent with observed incidence data, there are important differences in the projected courses of disease development and progression. For example, in all models, disease onset represents onset of biopsy-detectable tumors, so it is not surprising that projected risks of this event are lower than the estimated 36% of men with autopsy-detectable disease.[25] However, the PSA test sensitivity in the FHCRC model is lower than that in the MISCAN and UMICH models; in other words, onset in the FHCRC model represents onset of disease that is harder to detect than that in the other models. Consequently, this model projects higher probabilities of onset and, given onset, lower probabilities of clinical diagnosis and prostate cancer death.

Another example of conceptual differences in disease natural history carrying implications for projected summary measures is manifested in projected mean years from onset to clinical diagnosis. The sequential stage progression formulation in the MISCAN model implies longer durations for cancers that are metastatic by the time of clinical diagnosis (mean 13 years) than for all cancers (mean 9 years). In contrast, the UMICH model implicitly assumes that tumors that become metastatic by the time of clinical diagnosis tend to grow faster (mean duration to diagnosis 4 years) than all cancers (mean duration to diagnosis 7 years).

All models project an unconditional lifetime risk of clinical diagnosis that is slightly higher than the 9% estimated in Ries et al.[26] based on incidence trends prior to the advent of PSA screening. Lifetime risks of dying from prostate cancer are close to the 3% obtained from Devcan (version 7) software for the pre-PSA era.[26] Mean lead times associated with PSA screening range from 4 to 9 years and broadly agree with previously published estimates in the US setting.[5, 27-28]

The models also project risks of clinical diagnosis, metastasis by clinical diagnosis, and prostate cancer death in men aged 50–84 screen detected in SEER local-regional stage in 2000. Risks of these events are projected by age and grade for all three models and by age, grade, and PSA for the FHCRC model. As expected, risks fall with age (as non-prostate cancer death intervenes) and rise with more aggressive clinical characteristics.

It is worthwhile to observe three points related to these results. First, the risk that a PSA-detected patient would have gone on to be clinically diagnosed in the absence of screening is one minus the risk that he has been overdiagnosed. Thus, one minus the risk of clinical diagnosis reported in Tables 6.3 and 6.4 corresponds to the risk of overdiagnosis conditional on age and grade (and PSA) projected by the models. This complementary interpretation

may help patients to make treatment decisions tailored to personalized clinical information known at diagnosis rather than based on broad estimates of overdiagnosis in the population as have been published previously.[11]

Second, we find that even for men with Gleason score 2–7, the risks that the disease would lead to death in the absence of primary treatment (23–34% for men under age 60 at PSA detection) are nontrivial. Our estimates of disease-specific survival in this setting are based on men diagnosed in 1983–1986 without primary treatment, which we assume to be valid for contemporary cohorts. This assumption is imperfect because there have almost certainly been improvements in post-diagnosis monitoring of conservatively managed disease and a widened availability of more effective salvage therapies.[29] However, because there is great uncertainty about how to correct for these improvements over time, we present the results without adjustment. Because these therapeutic advances have likely improved prostate cancer survival, the projections reported here may underestimate present-day prostate cancer survival in the absence of primary treatment.

Even if men are not treated at the time of screen detection, it is likely that intervention will occur at some later date should clinical events warrant it. To determine how our projections of disease-specific deaths would change in this setting, we assumed that all screen-detected cases received radical prostatectomy at their date of clinical diagnosis, and, accordingly, we inflated their post-lead time survival by a factor of 0.65.[30] Under this assumption, the models projected that treatment reduced the risk of prostate cancer death by approximately 13–28% for men with Gleason score 2–7 and by approximately 13–20% for men with Gleason score 8–10. In absolute terms, for men under 60, the risk of prostate death decreased from 23–34% to 18–27% for Gleason score 2–7 disease and from 62–80% to 51–66% for Gleason score 8–10 disease.

These results have implications for management of PSA-detected prostate cancers. They suggest that, for younger men, the risks of progressing to lethal disease remain nontrivial even if curative treatment is pursued at clinical diagnosis. Since younger men are subject to low risks of overdiagnosis, these men would be well justified in considering primary treatment at PSA detection. However, we caution that the results presented here do not speak to the question of the optimal timing of primary therapy nor to the relative benefits of earlier versus later treatment. Instead, these results only provide alternative benchmark information for these treatment decisions. And as in other contexts, the optimal treatment decision should weigh any likely gains against known harms associated with treatments (for a review see[31]).

Third, the FHCRC model implies that the PSA level carries important information about the risks of clinical progression events, particularly for older men. For example, for men aged 65–69 detected with Gleason score 2–7, these risks are 58%, 8%, and 13% when PSA is 4–7 ng/mL at diagnosis and 80%, 11%, and 20% when PSA is over 10 ng/mL at diagnosis. Consistent with findings in other studies,[32] these results indicate that PSA information stratifies

risk. However, these results do not address whether early detection confers a survival benefit because any such benefit is contingent on treatment, which is disallowed by all models in this article. Despite this, the reader may be tempted to infer that PSA screening induces an implicit survival benefit because men with lower PSA at detection have a lower probability of dying of prostate cancer compared to men with higher PSA at detection. But these groups are not comparable due to selection artifacts. Specifically, the group of men with lower PSA at detection contains more overdiagnosed cases, and since overdiagnosed cases are not at risk of prostate cancer death, they artificially lower the risk of disease-specific mortality for this group as a whole. In addition, men with lower PSA at detection tend to have slower PSA growth in this model, which implies longer lead times and therefore also longer times to disease-specific mortality. This selection that occurs when a population is screened, and the artifacts that it generates, are key motivators for conducting randomized screening trials.

Randomized screening trials provide the opportunity to evaluate the benefits of screening in comparable groups and in the presence of treatment. Thus, these trials have best chance of identifying a survival benefit if one exists. The fact that the US and European prostate cancer screening trials did not show an unequivocal benefit[33-34] has led many to doubt whether such a benefit exists, but it has yet to be verified whether the conflicting results are due to lack of benefit or to differences in settings, protocols, and implementations.

Survival projections are relatively consistent across the models despite variability in the modeled rates of disease progression events. As for risks of prostate cancer death, it is important to note that the survival model is based on pre-PSA data and hence projections in the absence of primary treatment are also in the absence of present-day disease management practices. We compared our survival results with those of Nicholson and Harland[5] and Parker et al.[6] Projected 15-year prostate cancer survival rates appear to be consistent with Nicholson and Harland (Table 6.4 in their manuscript[5]); averages across the three CISNET models are 87% for men aged 50–54 to 92% for men aged 80–84 at PSA detection for Gleason scores 2–7, while they project 88% for men aged 50 to 96% for men aged 80 at PSA detection for Gleason score ≈ 6 (a complete list of age- and grade-specific projections could not be obtained with their model). In contrast, because Parker et al.[6] use grade categories 2–6, 7, and 8–10, we compare survival projections for Gleason score 8–10 only. We project survival probabilities of 50% for men aged 55–59 to 64% for men aged 70–74 at PSA detection, while they project 28% for men aged 55–59 to 72% for men aged 70–74 at PSA detection. Our projections, particularly those for younger men, may be more consistent with their alternative scenario which assumes shorter lead times and lower overdetection rates than were reported in Draisma et al.[8] These assumptions may be more realistic in the US setting.[11]

Finally, it is important to recognize that projected risks of clinical progression and disease-specific mortality for PSA-detected cases are based on the coarse categorization of Gleason scores 2–7 into a single category. This categorization was adopted to overcome the tendency of Gleason scores to migrate from well to moderately differentiated grades over calendar

time, which is difficult to model adequately. As a consequence, the model projections cannot distinguish cancers with Gleason scores 2–6 from the more aggressive cancers with Gleason scores 3+4 or 4+3. Therefore, the models may overstate the risks of clinical progression for men with lower Gleason scores in this category.

While there is broad qualitative agreement in the projections across models, there is also quantitative variability. This variability reflects real uncertainty about the natural history of the disease despite the successful calibration of each model to observed incidence data. Yet despite this uncertainty, the models provide important answers to questions that patients with PSA-detected prostate cancers consistently ask when faced with determining their optimal treatment course. The models provide estimates of the risks that their cancer has been overdiagnosed and that it would progress to a metastatic state or lead to death if left untreated. These risks are estimated based on factors that have been shown to be most predictive of prognosis, namely, age, grade, and PSA. If the projected risks of clinical progression outcomes are low, then patients can be reassured of the advisability of a conservative treatment approach. More generally, these tailored risks of disease progression events should be useful when selecting treatment, but it should be recognized that the extent to which treatment will alter these risks is not clear and is not addressed by the present manuscript.

Further work will examine how survival projections change if treatment is instituted some time after PSA detection but before clinical diagnosis, when there is evidence of disease progression. This future work will be contingent on obtaining reliable estimates for treatment efficacy conditional on its timing following PSA detection. Ongoing studies of active surveillance will hopefully provide us with this information, which will then be incorporated into the models to provide patients with even more tailored information about their likely outcomes under different treatment options.

REFERENCES

1. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
2. Johansson JE, Andren O, Andersson SO, et al. Natural history of early, localized prostate cancer. *Jama.* Jun 9 2004;291(22):2713-2719.
3. Albertsen P. Predicting survival for men with clinically localized prostate cancer: what do we need in contemporary practice? *Cancer.* Jan 1 2008;112(1):1-3.
4. Paquette EL, Sun L, Paquette LR, Connelly R, McLeod DG, Moul JW. Improved prostate cancer-specific survival and other disease parameters: impact of prostate-specific antigen testing. *Urology.* Nov 2002;60(5):756-759.
5. Nicholson PW, Harland SJ. Survival prospects after screen-detection of prostate cancer. *BJU Int.* Nov 2002;90(7):686-693.
6. Parker C, Muston D, Melia J, Moss S, Dearnaley D. A model of the natural history of screen-detected prostate cancer, and the effect of radical treatment on overall survival. *British Journal of Cancer.* May 22 2006;94(10):1361-1368.
7. Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin KA, Hankey BF. Cancer surveillance series: interpreting trends in prostate cancer--part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality. *J Natl Cancer Inst.* Jun 16 1999;91(12):1033-1039.
8. 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.* Jun 18 2003;95(12):868-878.
9. Tsodikov A, Szabo A, Wegelin J. A population model of prostate cancer incidence. *Stat Med.* Aug 30 2006;25(16):2846-2866.
10. Etzioni R, Gulati R, Falcon S, Penson DF. Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: a surveillance modeling approach. *Med Decis Making.* May-Jun 2008;28(3):323-331.
11. 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.* Mar 18 2009;101(6):374-383.
12. Surveillance, Epidemiology and End Results (SEER). http://seer.cancer.gov/.
13. Albertsen PC, Hanley JA, Barrows GH, et al. Prostate cancer and the Will Rogers phenomenon. *J Natl Cancer Inst.* Sep 7 2005;97(17):1248-1253.
14. 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.* May 1 2007;109(9):1877-1886.
15. National Cancer Institute, Cancer Intervention an Surveillance Modeling Network. http://cisnet.cancer.gov/.
16. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control.* Mar 2008;19(2):175-181.
17. 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.* Oct 2010;11(4):707-719.
18. Inoue LY, Etzioni R, Morrell C, Muller P. Modeling disease progression with longitudinal markers. *Journal of the American Statistical Association.* 2008;103(481):259-270.
19. Inoue LY, Etzioni R, Slate EH, Morrell C, Penson DF. Combining longitudinal studies of PSA. *Biostatistics.* Jul 2004;5(3):483-500.
20. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. *New England Journal of Medicine.* Jul 17 2003;349(3):215-224.
21. 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.* Mar 2005;173(3):746-750; discussion 750-741.
22. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
23. 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.* Mar 3 2010;102(5):352-355.
24. Chefo S, Tsodikov A. Stage-specific cancer incidence: an artificially mixed multinomial logit model. *Stat Med.* Jul 10 2009;28(15):2054-2076.
25. Etzioni R, Cha R, Feuer EJ, Davidov O. Asymptomatic incidence and duration of prostate cancer. *Am J Epidemiol.* Oct 15 1998;148(8):775-785.

26. Ries LAG, Harkins D, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2003, National Cancer Institute: http://seer.cancer.gov/csr/1975_2003/. In. Bethesda, MD; 2006.
27. Gann PH, Hennekens CH, Stampfer MJ. A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *Jama.* Jan 25 1995;273(4):289-294.
28. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics.* Mar 2008;64(1):10-19.
29. Klein EA, Platz EA, Thompson IM. Epidemiology, etiology, and prevention of prostate cancer. In: Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA, eds. Campbell-Walsh Urology. 9th ed. Philadelphia: Saunders. 2007:2854–2873.
30. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 12 2005;352(19):1977-1984.
31. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of Internal Medicine.* Mar 18 2008;148(6):435-448.
32. Kattan MW, Cuzick J, Fisher G, et al. Nomogram incorporating PSA level to predict cancer-specific survival for men with clinically localized prostate cancer managed without curative intent. *Cancer.* Jan 1 2008;112(1):69-74.
33. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
34. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.

Chapter 7

To be screened or not to be screened? Modeling the consequences of PSA screening for the individual

Elisabeth M.Wever, Jonas Hugosson, Eveline A.M. Heijnsdijk,
Chris H. Bangma, Gerrit Draisma and Harry J. de Koning

British Journal of Cancer. Epub ahead of print in July 2012. Doi: 10.1038/bjc.2012.317.
Reprinted with permission from Nature Publishing Group.

ABSTRACT

Background

Screening with prostate-specific antigen (PSA) can reduce prostate cancer mortality, but may advance diagnosis and treatment in time and lead to overdetection and overtreatment. We estimated benefits and adverse effects of PSA screening for individuals who are deciding whether or not to be screened.

Methods

Using a microsimulation model we estimated lifetime probabilities of prostate cancer diagnosis and death, overall life-expectancy and expected time to diagnosis, both with and without screening. We calculated anticipated loss in quality of life due to prostate cancer diagnosis and treatment that would be acceptable to decide in favor of screening.

Results

Men who were screened had a gain in life-expectancy of 0.08 years but their expected time to diagnosis decreased by 1.53 life-years. Of the screened men, 0.99% gained on average 8.08 life-years and for 17.43% expected time to diagnosis decreased by 8.78 life-years. These figures imply that the anticipated loss in quality of life due to diagnosis and treatment should not exceed 4.8%, for screening to have a positive effect on quality-adjusted life-expectancy.

Conclusions

The decision to be screened should depend on personal preferences. The negative impact of screening might be reduced by screening men who are more willing to accept the side-effects from treatment.

INTRODUCTION

The purpose of screening with prostate-specific antigen (PSA) is to reduce the risk of dying from prostate cancer by finding and treating prostate cancers at an early stage. The European Randomized study of Screening for Prostate Cancer (ERSPC) and the Göteborg trial (part of the ERSPC) showed that a reduction in prostate mortality can be obtained by screening with PSA.[1-2] However, an adverse effect of screening is overdiagnosis, i.e. the detection of cancers that would not have been diagnosed during the patients' lifetime if they had not been screened.[3] Also, all men with screen-detected prostate cancer have to live more years with the knowledge that they have prostate cancer (lead time). Men with screen-detected prostate cancer who opt for curative treatment risk living many years with the side-effects of treatment which would otherwise be symptom-free years.[4-5]

Most major U.S. medical organizations[6-8] and the European Association of Urology[9] recommend that clinicians discuss the potential benefits and adverse effects of PSA screening with their clients, consider their clients' preferences, and individualize screening decisions. Information on the consequences of PSA screening on a persons' life can help in making an informed decision.

This paper presents results of a simulation model showing the major differences between the potential course of lives of men who decide to be screened and those of men who decide not to be screened. We present the lifetime probability of prostate cancer diagnosis and death, overall life-expectancy and expected pre-diagnosis life-years (the average life-years till the time of prostate cancer diagnosis) for the two groups. Though several studies have presented benefits and adverse effects of PSA screening for a population,[1-3] we present the consequences of screening from the time of the decision to be screened or not. Also, while in empirical studies harms and benefits are calculated after a follow-up time of some years, this study presents results for the whole lifetime.

Health economists use the concept of utility to express quantitatively adverse effects of diagnosis and treatment on quality of life. In this view life without prostate cancer has utility 1; after diagnosis and treatment utility decreases to a lower level, depending on individual preferences and the specific consequences of diagnosis and treatment. Several authors have obtained estimates of utility or quality of life of living with these consequences.[10-13] While we do not estimate the utility and quality of life of living with prostate cancer, our results do allow calculation of the utility level below which the expected harms of early diagnosis exceed the expected gains from prevented prostate cancer mortality (the break-even point), which can be compared with individual expectations.

METHODS

We used the MISCAN (MIcrosimulation SCreening ANalysis) prostate cancer model. MISCAN is a micro-simulation program that simulates the progression and screening of prostate cancer. The model was validated on prostate cancer detection data from the ERSPC Rotterdam[14-16] and the ERSPC Göteborg,[17] and on the mortality reduction data from the overall ERSPC trial.[2, 18] The assumptions of the model and the estimation of the parameters are explained in the appendix. For this analysis a cohort model was used, with the age of first screening distributed uniformly from age 50 to 70 and with subsequent screening until age 75. The lifetime risks of prostate cancer diagnosis and death were calculated for the situation in which there was no screening and that with annual screening.

MISCAN model

A detailed description of the model and the data sources that informed the quantification of the model can be found in previous studies[14-16, 18] and also in a standardized model profile: http://cisnet.cancer.gov/prostate/profiles.html. MISCAN is a micro-simulation program that simulates the progression of prostate cancer in individuals as a sequence of preclinical, clinical and screen-detected tumor states. First, the age at death from other causes is simulated per individual using Dutch life tables (Statistics Netherlands, 2000-2007). Next, the progression of prostate cancer in the absence of screening is simulated. Prostate cancer may develop from no prostate cancer to a clinically diagnosed cancer through one or more screen-detectable preclinical stages. From each preclinical stage, a tumor may grow to the next clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; T3+, palpable, with extensions beyond the prostatic capsule); it may dedifferentiate to a higher Gleason score (well differentiated, Gleason score 2-6; moderately differentiated, Gleason score 7; poorly differentiated, Gleason score 8-10); or it may be clinically diagnosed. For these transitions, the time spent in the current stage is generated from a Weibull distribution, where the parameters depend on the current stage and the choice of next stage is determined by transition probabilities. In addition, there is a risk that a tumor in the local-regional stage (M0) will develop into disseminated disease (M1), which is modeled by using a stage and Gleason score-specific hazard function. Depending on the frequency and sensitivity of the screening test, preclinical cancers may be detected by screening. PSA test and subsequent biopsy were modeled as a single test, where the sensitivity parameter was assumed to be clinical T-stage dependent. In the model, sensitivity is defined as the probability that a preclinical tumor is detected by a screening test at the time the test is taken.

Model parameters, including transition probabilities, mean dwelling times (the time from one preclinical state to another preclinical or clinical state), and stage-specific test sensitivities were estimated by constructing models for the ERSPC-Rotterdam and Göteborg, and by calibrating the model to the following data observed at these centers: baseline incidence (Na-

tional Cancer Registry data for 1991)[19] and stage distribution in the Netherlands (Rotterdam cancer registry data 1992-1993);[20] baseline incidence in Sweden (1988-1992[21]); incidence, Gleason and stage distributions in the control arms of ERSPC-Rotterdam and Göteborg; and detection rates, interval cancer rates, Gleason and stage distributions in the screen arms of ERSPC-Rotterdam[14-16] and Göteborg.[17] Number of cases diagnosed, and Gleason and stage distributions in the control arms versus those in the screen arms provide insight into disease progression through the various preclinical phases. Parameters were estimated by numerically minimizing the deviance between the number of cases observed and the number of cases predicted by the models. Deviances were calculated by assuming Poisson likelihood for incidence data and by assuming multinomial likelihood for stage-distribution data. The parameters for incidence, clinical diagnosis and sensitivity were assumed to be country-specific. For the main results in this analysis we used the country-specific parameters corresponding to the Netherlands. The cohort models used for the analysis were constructed by changing the screening protocol of the ERSPC-Rotterdam model to the screening protocols considered in this study.

We assumed that if men are clinically diagnosed with prostate cancer, the time to death from prostate cancer is determined by prostate cancer-specific survival curves. Survival curves for men with no initial treatment were estimated on the basis of SEER (Surveillance, Epidemiology and End Results) data, specifically on cases diagnosed in the pre-PSA era (1983-1986). We assumed that all men diagnosed with prostate cancer receive radical prostatectomy. According to the published results,[22] we assumed that men receiving radical prostatectomy have a relative risk of 0.62 of dying from prostate cancer compared to men receiving no initial treatment. This analysis did not consider other treatments, such as radiation therapy and active surveillance, because of the limited published results about the effectiveness of these treatments. If the effectiveness of radiation therapy and active surveillance are different than that of radical prostatectomy, the results for these treatments will be different than those presented. For distant prostate cancer, we assumed that treatment has no effect. Some treatments might increase survival of distant disease slightly. However, in our view this is a minor limitation of the model.

The effect of early detection through screening on survival was included by assuming that a fraction of local-regional tumors detected by screening are cured because the tumors are treated earlier. The Gleason score-dependent cure rates were estimated by calibrating the ERSPC-Rotterdam model to the observed 27% prostate cancer mortality reduction in the overall ERSPC at follow-up after nine years.[18]

Calculating utility break-even point

We assigned being alive without diagnosed prostate cancer a utility of one, being dead a utility of zero, and being alive with diagnosed and treated prostate cancer a utility of *u*. The utility break-even point is the value of the utility of living with diagnosed and treated prostate

cancer, *u*, for which the expected quality of life lost due to earlier detection and treatment of cancer is equal to the expected quality adjusted life-expectancy gained (Figure 7.1). Therefore, men who can judge that their reduction in quality of life in the event that they are diagnosed and treated for prostate cancer exceeds the reduction that is represented by this threshold utility should possibly refrain from screening participation.

Sensitivity analyses

To evaluate the statistical variation and uncertainty of the observed data we conducted a number of sensitivity analyses. The sensitivity analyses compared models with various lead times, incidence, survival curves and cure rates. Penalized optimization was used to obtain a range of models with various lead times. Parameters for these models were estimated by minimizing the sum of total deviance and lead time penalty (mean lead time x penalty). The penalties used were -100 (favoring long lead times), 0 and 100 (favoring short lead times). We varied incidence in the models by using estimated incidence parameters that reproduced incidence in the Netherlands or Sweden. Different survivals were considered by assuming a relative risk of 0.5, 1 or 2 on the hazard of prostate cancer death. The various cure rates were obtained by calibrating the ERSPC-Rotterdam model to a mortality reduction of 27%

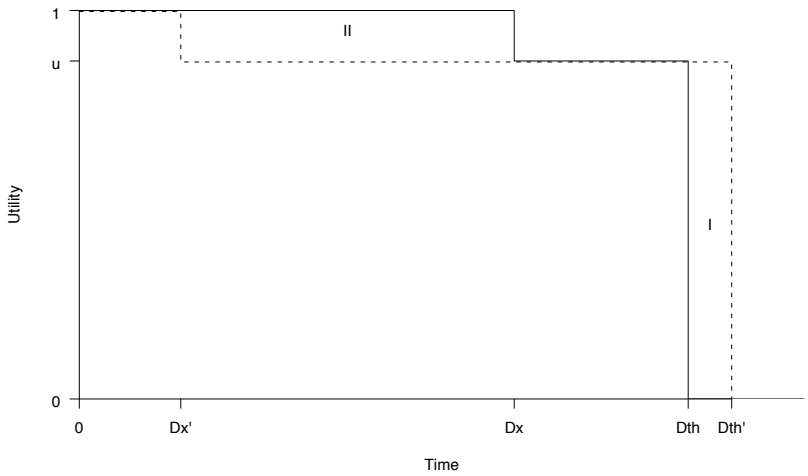


Figure 7.1: Harms and benefits in prostate cancer screening. Time 0 is the time of deciding to participate or not in screening. Utility or quality of life has value 1 until the moment of a prostate cancer diagnosis (Dx); the remaining lifetime until death (Dth) has utility u < 1. The figure shows hypothetical utility curves for a person without (solid) and with screening (dashes). Screening may detect prostate cancer earlier (at time (Dx') and possibly postpone the moment of death (to time Dth'). Area I represents the gain in quality adjusted life-years, and area II the loss in quality of life due to earlier detection. The level u depends on the consequences of diagnosis and treatment. If the *expected* gain in quality adjusted life-years (area I) equals the *expected* loss of quality of life due to earlier diagnosis (area II), the decision to participate in screening or not does not affect expected quality adjusted life-years. The utility break-even point is the utility level corresponding to that situation.

(observed for attendees in ERSPC), 44% (observed for screen-group in ERSPC-Göteborg) or 56% (observed for attendees in ERSPC-Göteborg).

RESULTS

Table 7.1 shows predicted risks of prostate cancer diagnosis and death and life-expectancies for men aged 50-54, 55-59, 60-64 and 65-69, both when deciding to participate in screening or not. As an illustration, consider men in the 55-59 age-group. Men who had not been screened had a 14.44% lifetime risk of being diagnosed with prostate cancer and a 2.94% lifetime risk of dying from it. Life-expectancy was 22.31 years, 21.08 of which pre-diagnosis. By comparison, those who participated in a yearly screening program had a 21.54% lifetime risk of being diagnosed with prostate cancer, a 1.89% risk of dying from prostate cancer and

Table 7.1: Predicted results of prostate cancer (pc) diagnosis and death.

Scenario	Lifetime probability of pc diagnosis (%)	Lifetime probability of pc screen-detection (%)	Lifetime probability of pc death (%)	Pre-diagnosis life-years*	Post-diagnosis life-years	Life-expectancy (years)	Utility break-even point‡
Age at first screen 50-54							
No screening	14.44	0.00	2.98	25.44	1.28	26.72	-
Sscreening†	21.38	17.62	1.91	23.81	2.99	26.81	0.947
Age at first screen 55-59							
No screening	14.44	0.00	2.94	21.08	1.24	22.31	-
Screening	21.54	17.73	1.89	19.47	2.93	22.40	0.949
Age at first screen 60-64							
No screening	14.07	0.00	2.86	17.07	1.11	18.18	-
Screening	21.38	17.51	1.88	15.55	2.71	18.26	0.954
Age at first screen 65-69							
No screening	13.35	0.00	2.64	13.50	0.93	14.43	-
Screening	20.90	16.79	1.81	12.15	2.34	14.49	0.960
Age at first screen 50-69 (All)							
No screening	14.10	0.00	2.86	19.55	1.15	20.70	
Screening	21.31	17.43	1.87	18.02	2.76	20.78	0.952

*Expected life-years till the time of prostate cancer diagnosis.
†Annual screening.
‡The utility break-even point is the value of the utility of living with diagnosed prostate cancer for which the utility-adjusted life-expectancy does not change upon deciding to participate in screening or not. Its value decreases with larger gains in overall life-expectancy relative to the expected loss in pre-diagnosis life-years. A high value of the utility break-even point means that men should only decide in favor of screening when they anticipate a small loss in quality of life due to detection and possibly treatment of prostate cancer.

a life-expectancy of 22.40 years, 19.47 of which pre-diagnosis. Therefore, men screened annually increased their life-expectancy by 0.09 years, i.e. 33 days. However, because of the lead time inherent in screen detection their expected pre-diagnosis life-years decreased by 1.61 years, i.e. 588 days.

Though the 0.09 expected gain in life-years appears small, this is a weighted average of all men who were screened. Among men who were screened, a fraction of 1.05% (=2.94-1.89%) did not die of prostate cancer because the cancer was detected and treated earlier (Table 7.2). Life-expectancy among these men increased by 8.57 years (=0.09*100/1.05). The expected loss of 1.61 pre-diagnosis life-years is also a weighted average. Among men who were screened, 17.73% were screen-detected with prostate cancer. Their expected pre-diagnosis life-years decreased by 9.08 years (=1.61*100/17.73).

Figure 7.2 presents the probability of being alive without diagnosed prostate cancer, alive with diagnosed prostate cancer, dead from prostate cancer, and dead from other causes at various points in time, from the time of the decision to be screened or not, and how these probabilities were affected by the decision. Patients in the health state alive with diagnosed prostate could be cured of their disease or not. This health state is a transient state that eventually will be absorbed by the state dead from prostate cancer or dead from other causes. The figures show that although screening reduced the probability of dying from prostate cancer in the long term, it substantially increased the probability of being alive with diagnosed prostate cancer in the short term. This pattern is especially clear in older ages, who have a higher probability of having a preclinical prostate cancer, which is screen-detectable. For younger ages, too, screening substantially increased the probability of being alive with diagnosed prostate cancer, though in their case this increase was more widely distributed over the follow-up years.

Table 7.1 and Figure 7.2 can be used by clinicians and individuals to discuss the short and long term benefits and harms that apply to the average individual. To summarize the harms

Table 7.2: Percentage of population experiencing benefits and harms of prostate cancer (pc) screening. Results are for men aged 55-59 at first screen.

	% Of population	Decrease in pre-diagnosis life-years	Increase in life-expectancy (years)
Never diagnosed with pc whether being screened or not	78.46	0.00	0.00
Diagnosed with pc if being screened, but never diagnosed with pc if not being screened	7.10	10.69	0.00
Diagnosed with pc and dead from other causes whether being screened or not	11.50	5.43	0.00
Diagnosed with pc and dead from pc whether being screened or not	1.89	6.97	0.00
Diagnosed with pc whether being screened or not, but not dead from pc because of early detection by screening	1.05	9.15	8.57
Weighted total	100.00	1.61	0.09

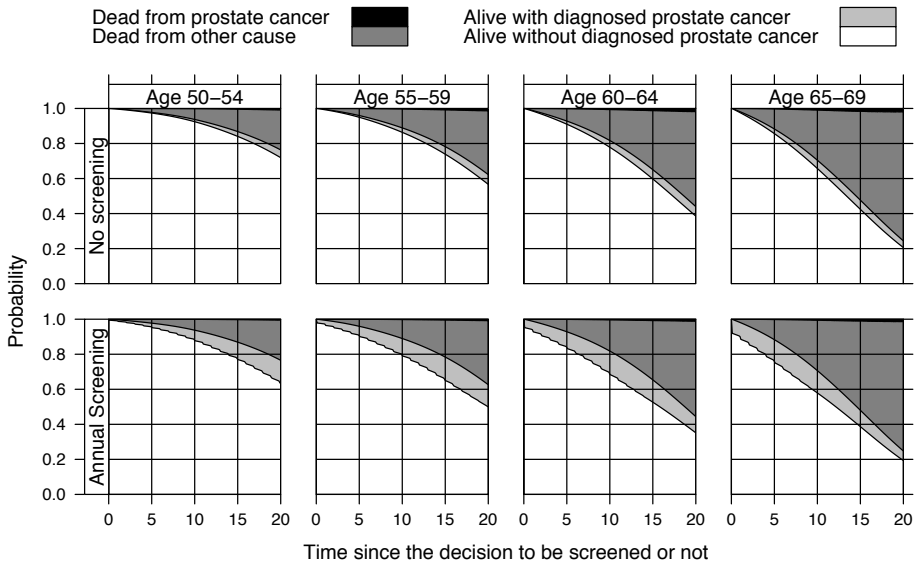


Figure 7.2: Survival curves with follow-up time from the time of decision. These stacked figures show the proportion of men who are alive without diagnosed prostate cancer (white area), alive with diagnosed prostate cancer (light grey area), dead from prostate cancer (dark grey area) and dead from other causes (black area) at various points in time.

and benefits in one measure we calculated the utility break-even point. The utility break-even points were high (0.947-0.960) for all ages. These utility break-even points imply that men who might judge, based on information given, that their quality of life will decrease by more than 4.0-5.3% in the event that they are diagnosed and treated for prostate cancer should probably avoid being screened. It is important to note that the utility break-even points are estimated clinical outcomes from a population that applies to average individuals.

Varying lead time, incidence, survival and cure rates caused the decrease in expected pre-diagnosis life-years to vary from 1.40 to 2.31 years, the increase in life-expectancy to vary from 0.02 to 0.40 years and the utility break-even point to vary from 0.833 to 0.991 (Table 7.3). Shorter lead times, more negative survival and higher cure rates yielded results more in favor of screening. While the loss in quality of life acceptable for a man to consider screening was only 0.9% in the most unfavorable model for screening, the most favorable model showed a 16.7% loss as still being acceptable.

DISCUSSION

The analysis showed that if an individual decides to be screened, his lifetime risk of prostate cancer death decreased from 2.86% to 1.87%, while his overall life-expectancy increased by 0.08 years. At the same time, his lifetime risk of being diagnosed with prostate cancer

Table 7.3: Sensitivity analysis for uncertainty in the model and the data.*

Mean lead time†	Incidence‡	Relative risk on pc death§	Cure rate‖	Scenario	Lifetime probability of pc diagnosis (%)	Lifetime probability of pc death (%)	Pre-diagnosis life-years*	Post-diagnosis life-years	Life-expectancy (years)	Utility break-even point¶
8.02	Rotterdam	1	0.42 and 0.22	No screening	14.44	2.94	21.08	1.24	22.31	-
				Annual screening	21.54	1.89	19.47	2.93	22.40	0.949
Varying lead time										
8.50	Rotterdam	1	0.42 and 0.22	No screening	13.14	2.68	21.22	1.11	22.34	-
				Annual screening	21.47	1.73	19.45	2.97	22.41	0.959
7.12	Rotterdam	1	0.42 and 0.22	No screening	15.39	3.03	20.95	1.35	22.30	-
				Annual screening	21.21	1.95	19.55	2.84	22.39	0.939
Varying incidence										
8.02	Göteborg	1	0.42 and 0.22	No screening	20.70	4.23	20.44	1.77	22.21	-
				Annual screening	30.87	2.72	18.13	4.20	22.34	0.949
Varying survival										
8.02	Rotterdam	0.5	0.42 and 0.22	No screening	14.44	1.78	21.08	1.33	22.41	-
				Annual screening	21.54	1.14	19.47	2.99	22.46	0.969
8.02	Rotterdam	2	0.42 and 0.22	No screening	14.44	4.48	21.08	1.11	22.19	-
				Annual screening	21.54	2.91	19.47	2.85	22.32	0.924

Varying cure rate										
8.02	Rotterdam	1	0.67 and 0.35	No screening	14.44	2.94	21.08	1.24	22.31	-
				Annual screening	21.54	2.56	19.47	2.88	22.35	0.981
8.02	Rotterdam	1	0.86 and 0.44	No screening	14.44	2.94	21.08	1.24	22.31	-
				Annual screening	21.54	0.82	19.47	3.02	22.49	0.902
Most unfavorable model for screening										
8.50	Rotterdam	0.5	0.42 and 0.22	No screening	13.14	1.62	21.22	1.20	22.42	-
				Annual screening	21.47	1.41	19.45	2.99	22.44	0.991
Most favorable model for screening										
7.12	Göteborg	2	0.86 and 0.44	No screening	21.84	6.62	20.29	1.72	22.01	-
				Annual screening	30.10	1.97	18.30	4.11	22.41	0.833

*Results are presented for men who were first screened at age 55-59.

†We used penalized optimization to obtain a range of models with different lead times (Draisma et al., 2003).

‡We varied the incidence in the model by using the estimated incidence parameters that reproduced the incidence in the Netherlands or Sweden.

§A relative risk of 0.5 on the hazard of prostate cancer death increases prostate cancer-specific survival and a relative risk of 2 on the hazard of prostate cancer death decreases prostate cancer-specific survival.

||The cure rates show the proportion of screen-detected men who do not die of pc because their cancer is diagnosed and treated earlier. The cure rates were obtained by calibrating the ERSPC-Rotterdam model to a 10%, 27% or 56% mortality reduction. The first cure rates are for men screen-detected for pc with Gleason score <=7 and the second cure rates are for men screen-detected for pc with Gleason score >7.

¶The utility break-even point is the value of the utility of living with diagnosed prostate cancer for which the utility-adjusted life-expectancy does not change upon deciding to participate in screening or not. Its value decreases with larger gains in overall life-expectancy relative to the expected loss in prostate-cancer-free life-expectancy. A high value of the utility break-even point means that men should only decide in favor of screening when they anticipate a small loss in quality of life due to detection and possibly treatment of prostate cancer.

increased from 14.10% to 21.31% and his expected pre-diagnosis life-years decreased by 1.53 years. A fractional percentage of 0.99% of screened men enjoyed a benefit of living an average of 8.08 years longer, while 17.43% of the screened men were screen-detected and lived an average of 8.78 pre-diagnosis life-years less. In many cases this means an increase in life-years suffering from morbidity associated with treatment.[5, 23]

Albertsen et al. estimated risks of death from prostate cancer as a function of grade and age for prostate cancer left untreated (not treated curatively).[24-25] His curves show that the risk of dying from prostate cancer is high for high grade cancer only. His estimates are relevant for men diagnosed with prostate cancer who have to decide on treatment. Similarly, our Figure 7.2 provides information that could be relevant for men deciding to be screened or not. It shows that screening reduces a small risk of dying from prostate cancer in the future, at the cost of substantial increased risk of being diagnosed with prostate cancer in the short term.

The high utility break-even points (0.947-0.960) in this analysis imply that only a small anticipated loss in quality of life after diagnosing and treating prostate cancer would be acceptable for deciding in favor of screening. Because individuals assess the side-effects of treatment very differently[12, 26] the anticipated loss in quality of life might vary considerably between individuals. Some patients might find it very important to be potent or continent, for example, because they have an active life which they really want to maintain. Assume a scale from 0 to 100, where 100 implies that the individual would not mind at all to have sexual problems and 0 implies that the individual would mind very much. An individual whose value is 70 has an expected utility of 0.75 (= 1*12/(100-31)+0.70*57/(100-31)) in the event that he is diagnosed and treated for prostate cancer and therefore an anticipated loss in quality of life of 25%. This calculation is based on the results of Korfage et al.,[5] who presented that 31% of men had erectile dysfunction before radical prostatectomy and 88% after radical prostatectomy. In this case the anticipated loss in quality of life after a prostate cancer diagnosis and treatment will be higher than the utility break-even point and therefore the individual should refrain from screening participation. Other patients might find it less important to be potent or continent, which could mean that the anticipated loss in quality of life is lower than the utility break-even point and that therefore the individual should decide to be screened.

Considering the utility break-even point one should however acknowledge that expected utility theory is not always a good predictor of patient's actual decisions. Just as with insurance against fire, where the costs are much higher than the expected benefits, people may deem the adverse effects to be acceptable in exchange for the benefits received, while this might not be the case according to expected utility theory. Also, while it is already difficult to assess the loss in quality of life for individuals who are diagnosed and treated for prostate cancer,[27] it is probably even more difficult to assess the anticipated loss in quality of life for individuals who are not even screened yet. Therefore, it might be difficult to use the utility break-even points. To be able to use the utility break-even point a formal decision aid should

be constructed, which explains the possible side-effects of diagnosis and treatment, and which can provide a value for the anticipated loss in quality of life for each individual. For this decision aid published data[5, 10, 13] for the risks of the potential side-effects of treatment of prostate cancer can be used.

Another limitation of our study is that the results were calculated using data from a specific population, namely the ERSPC Rotterdam and Göteborg. Results may be different for other populations, since different incidences of and mortalities for prostate cancer have been observed in different countries.[28] The sensitivity analyses give an indication of the expected impact on the results for populations with different incidence, survival and cure rates. In Sweden, prostate cancer accounts for 5.5% of all causes of deaths among men[29] and the mortality reduction observed in the Göteborg-trial was 56%, among attendees; the expected impact for the Swedish population is therefore comparable to the results of the most favorable model for screening. In the U.S., the estimated lifetime risk of prostate cancer diagnosis is 16.22% and of prostate cancer death is 2.79%,[30] which are close to the main results presented. Therefore, the main results might also be applied in the U.S. population.

The results of this analysis apply to the average population. In this study we did not do sub-group analyses. However, there might be factors, for example, family history with prostate cancer, obesity and African-American race that imply an increased risk of prostate cancer. Men with these risk factors might have different results for the benefits and harms.

The third limitation is that we used survival curves based on data observed in the U.S. for the ERSPC-Rotterdam population modeled. We used these survival data since it is one of the few datasets presenting survival of untreated prostate cancer as a function of age at diagnosis and Gleason score progression. However, for our main results we estimated a 2.86% lifetime risk of prostate cancer death, which corresponds well with US[30] and Dutch data.[31]

The fourth limitation is that we did not consider the different treatments available, since there are limited published results about the effectiveness of all treatments. For the purposes of this analysis, we assumed that all individuals diagnosed with prostate cancer received radical prostatectomy. If the effectiveness of active surveillance, radiation therapy or other treatments (e.g. proton beam and brachytherapy) are different than that of radical prostatectomy, the benefits for these treatments will be different than those presented. Also, these might have different potential side-effects, which should be taken into account when determining the anticipated loss in quality of life.

If active surveillance is considered an option, treatment might be delayed. This implies that the benefit of screening might be reduced since immediate curative treatment should be at least as effective as delayed curative treatment. In the sensitivity analysis we considered a scenario where the benefit of screening is reduced (assuming a prostate cancer mortality reduction of 10% instead of 27% for screened men). The results for this scenario show the impact on the benefit of screening (0.03 versus 0.09 life-years gained, Table 7.2) if screening followed by active surveillance would achieve a mortality reduction of 10%. However,

since the effectiveness of active surveillance is not known it is not possible to calculate the exact benefits. Since with active surveillance the side-effects of treatment are postponed the anticipated loss in quality of life will also be lower for the individual.

To evaluate the statistical variation and the uncertainty of the observed data we performed sensitivity analyses. The models with different lead times show the statistical variation in the model as we used penalized optimization to obtain the parameters for these models. The models with different survival curves show the results for populations with worse or better prostate cancer survival. Since in the different centers of the ERSPC different prostate cancer mortality reductions has been observed we considered different cure rate estimates. The analysis shows that in populations with worse prostate cancer prognosis and higher cure rates 16.7% loss in quality of life is still acceptable for an individual to decide in favor of screening, which is quite higher than the 5.1% in the basecase.

After a median follow-up of nine years, the ERSPC-trial stated that the prevention of one prostate cancer death entailed the screening of 1,410 men (NNS: number needed to screen), and the additional diagnosis of 48 men (NNT: number needed to treat).[2] At 14 years of follow-up, the Göteborg-trial showed that the NNS was 293, and the NNT 12.[1] Welch et al. showed that, in the U.S. population in the 1986-2005 period, the NNT was maximally 23.[3] Using the lifetime risk of prostate cancer diagnosis and death in this study, the average NNS was 101.01 (=100/(2.86-1.87)) and the average NNT was 7.28 (=(21.31-14.10)/(2.86-1.87)). This illustrates the importance of complete follow-up for correctly estimating these quantities.

In conclusion, the results of the present analysis show that individuals who decide to be screened may lower their risk of dying from prostate cancer and live longer, but also that the associated adverse effects can be significant. However, the magnitude of the adverse effects depends on how undesirable it is for the individual to live with diagnosed and treated prostate cancer and the potential side-effects. The negative impact of screening might be reduced by screening men who are more willing to accept the side-effects from treatment.

REFERENCES

1. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *The lancet oncology.* Aug 2010;11(8):725-732.
2. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
3. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986-2005. *J Natl Cancer Inst.* Oct 7 2009;101(19):1325-1329.
4. Heijnsdijk EA, der Kinderen A, Wever EM, Draisma G, Roobol MJ, de Koning HJ. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British Journal of Cancer.* Dec 1 2009;101(11):1833-1838.
5. 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.* Aug 20 2005;116(2):291-296.
6. Lim LS, Sherin K. Screening for prostate cancer in U.S. men ACPM position statement on preventive practice. *American journal of preventive medicine.* Feb 2008;34(2):164-170.
7. Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW. Cancer screening in the United States, 2010: a review of current American Cancer Society guidelines and issues in cancer screening. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):99-119.
8. Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):70-98.
9. Heidenreich A, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. *European Urology.* Jan 2011;59(1):61-71.
10. Carlsson S, Aus G, Bergdahl S, et al. The excess burden of side-effects from treatment in men allocated to screening for prostate cancer. The Goteborg randomised population-based prostate cancer screening trial. *Eur J Cancer.* Mar 2011;47(4):545-553.
11. Dale W, Basu A, Elstein A, Meltzer D. Predicting utility ratings for joint health States from single health States in prostate cancer: empirical testing of 3 alternative theories. *Med Decis Making.* Jan-Feb 2008;28(1):102-112.
12. Steineck G, Helgesen F, Adolfsson J, et al. Quality of life after radical prostatectomy or watchful waiting. *New England Journal of Medicine.* Sep 12 2002;347(11):790-796.
13. Stewart ST, Lenert L, Bhatnagar V, Kaplan RM. Utilities for prostate cancer health states in men aged 60 and older. *Medical care.* Apr 2005;43(4):347-355.
14. 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.* Jun 18 2003;95(12):868-878.
15. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
16. 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.* Mar 3 2010;102(5):352-355.
17. Hugosson J, Aus G, Lilja H, Lodding P, Pihl CG. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. *Cancer.* Apr 1 2004;100(7):1397-1405.
18. Wever EM, Draisma G, Heijnsdijk EA, de Koning HJ. How does early detection by screening affect disease progression?: modeling estimated benefits in prostate cancer screening. *Med Decis Making.* Jul-Aug 2011;31(4):550-558.
19. Visser O, Coeberg JW, Schouten LJ. Incidence of cancer in the Netherlands,1991. *Utrecht (The Netherlands): The Netherlands Cancer Registry.* 1994:13, 45.
20. Spapen SJ, Damhuis RA, Kirkels WJ. Trends in the curative treatment of localized prostate cancer after the introduction of prostate-specific antigen: data from the Rotterdam Cancer Registry. *BJU Int.* Mar 2000;85(4):474-480.
21. Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. *Cancer Incidence in Five Continents, Vol. VII* Lyon: IARC;1997.
22. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.
23. Potosky AL, Legler J, Albertsen PC, et al. Health outcomes after prostatectomy or radiotherapy for prostate cancer: results from the Prostate Cancer Outcomes Study. *J Natl Cancer Inst.* Oct 4 2000;92(19):1582-1592.

24. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
25. Albertsen PC, Moore DF, Shih W, Lin Y, Li H, Lu-Yao GL. Impact of co-morbidity on survival among men with localized prostate cancer. *J Clin Oncol.* Apr 1 2011;29(10):1335-1341.
26. Korfage IJ, Hak T, de Koning HJ, Essink-Bot ML. Patients' perceptions of the side-effects of prostate cancer treatment--a qualitative interview study. *Social science & medicine (1982).* Aug 2006;63(4):911-919.
27. Essink-Bot ML, Korfage IJ, De Koning HJ. Including the quality-of-life effects in the evaluation of prostate cancer screening: expert opinions revisited? *BJU Int.* Dec 2003;92 Suppl 2:101-105.
28. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin D. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 2010; http://globocan.iarc.fr. Accessed December 28, 2011.
29. Welfare NBoHa. Official Statistics of Sweden. Causes of death 2007 http://www.socialstyrelsen.se/publikationer2009/dodsorsaker2007 (accessed April 29, 2011). 2009.
30. SEER. *Cancer Statistics Review 1975-2007*: National Cancer Institute; April: accessed at http://seer.cancer.gov/csr/1975_2007/results_merged/topic_lifetime_risk_diagnosis.pdf on March 23, 2011 2010.
31. CBS. Centrale Bureau van Statistieken. http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=03799&D1=291&D2=0&D3=0&D4=1-9&HD=120201-1423&HDR=T&STB=G1,G2,G3&CHARTTYPE=1 (accessed February 01, 2012). 2010.

Chapter 8

Quality of life effects of prostate-specific antigen screening

Eveline A.M. Heijnsdijk, Elisabeth M. Wever, Anssi Auvinen, Jonas Hugosson, Stefano Ciatto, Vera Nelen, Maciej Kwiatkowski, Arnauld Villers, Alvaro Páez, Sue M. Moss, Marco Zappa, Teuvo LJ Tammela, Tuukka Mäkinen, Sigrid Carlsson, Ida J. Korfage, Marie-Louise Essink-Bot, Suzie J. Otto, Gerrit Draisma, Chris H. Bangma, Monique J. Roobol, Fritz H. Schröder and Harry J. de Koning

New England Journal of Medicine. 2012;367(7):595-605.
Reprinted with permission from Massachusetts Medical Society.

ABSTRACT

Background

The European Randomized Study of Screening for Prostate Cancer (ERSPC) reported a 29% prostate cancer mortality reduction among screened men after 11 years. However, it is uncertain to what extent harms from overdiagnosis and treatment on quality of life counterbalance this benefit.

Methods

Based on ERSPC follow-up data, we used micro-simulation modeling (MISCAN) to predict the number of prostate cancers, treatments, deaths and quality-adjusted life-years (QALYs) gained following the introduction of screening. Various screening strategies, efficacies, and quality of life assumptions were modeled.

Results

Per 1,000 men of all ages followed for their entire lifespan we predicted for annual screening from age 55-69 years: 9 fewer deaths due to prostate cancer (28% reduction, 37% among screened men), 14 fewer men receiving palliative therapy (35% reduction), and 73 life-years gained (average 8.4 years per prostate cancer death avoided). QALYs gained were 56 (range: -21, 97), a reduction of 23% from unadjusted life-years gained (21% using 4-years intervals). The number needed to screen (NNS) was 98 and number needed to detect (NND) 5. Also inviting men aged 70-74 resulted in more life-years (82) but similar QALYs (56).

Conclusions

Although NNS and NND are more favorable than previously calculated, the benefit of PSA screening is diminished by loss of QALYs, that is dependent primarily on post-diagnosis long-term utility assumptions. Longer follow-up data from both the ERSPC and quality of life are essential before making universal recommendations regarding screening.

INTRODUCTION

The initial results of the European Randomized Study of Screening for Prostate Cancer (ERSPC) showed a significant prostate cancer mortality reduction in the screening group of 20% after a median follow-up of nine years, and of 27% in screened men when adjusted for selection bias.[1] The results have recently been updated, resulting after 11 years in a prostate cancer mortality reduction of 29% in men screened when adjusted for selection bias.[2] The Gothenburg trial, one center of the ERSPC, reported a prostate cancer mortality reduction of 44% after a median follow-up of 14 years and a 56% reduction for men screened at least once.[3] The PLCO trial found no mortality reduction in the screening group, however the rate of contamination was high and biopsy compliance low.[4]

Prostate cancer mortality reduction, life-years gained and a reduction of advanced disease are obvious benefits of screening. However, PSA screening is associated with considerable unfavorable effects. In the ERSPC screening group, the cumulative incidence of prostate cancer was 7.4%, versus 5.1% in the control group.[2] A proportion of the screen-detected tumors (10-56%) would never have led to clinical symptoms[5-8] but these overdiagnosed cancers are frequently treated nonetheless with associated risks of adverse effects.[9] Furthermore, because of a long lead time, estimated at 5-12 years,[6, 10] men have to live longer with those effects.

Reports on the harms and benefits of PSA screening are highly inconsistent due to the lack of results from randomized screening trials.[11-12] However, as more mature data from the ERSPC are available, for the first time realistic predictions of the effects of screening can be made. Therefore, this study quantifies the effects of screening strategies on prostate cancer mortality and quality of life, using a model based on data from the ERSPC. In addition, we have determined the harms and benefits for a range of treatment impact and mortality reduction scenarios.

METHODS

ERSPC data

The ERSPC was initiated in the early 1990s to evaluate the effect of PSA screening on prostate cancer mortality.[13] In seven countries, 162,243 men were randomized. Most centers used a PSA cutoff value of 3.0 ng/mL as an indication for biopsy, others used 4.0 ng/mL, with additional tests for values between 2.5 and 4.0. The screening interval was 4 years, with the exception of Sweden (2 years). Treatment was performed according to local policies and guidelines, independent of trial arm.[14] In line with the protocol, the effect of screening in the core age group (55-69 years) was evaluated. Follow-up data on mortality until 31 December 2008 are currently available.[2]

We used Microsimulation Screening Analysis (MISCAN) to extrapolate the results to alternative screening strategies and an extended follow-up.

Screening strategies

A population of men aged 0-100 years was simulated with an age distribution according to the European Standard Population.[15] The following screening strategies were simulated: annual screening in the age groups 55-69 years and 55-74 years, screening at 4-year intervals between 55-69, and single screens performed either at age 55, 60 or 65 years. An 80% participation proportion was assumed.

Quality of life

Quality-adjusted life-years (QALYs) were predicted using utility estimates for various health states. The utility estimates were obtained from the CEA Registry[16] and literature (Table 8.1) and ranged from 0 (death or worst imaginable health) to 1 (full health). In addition, data from ERSPC on treatment-related complications as urinary incontinence, bowel dysfunction and erectile dysfunction were analyzed. Favorable and unfavorable values were assigned according to the minimum and maximum values in the cited references. A utility estimate of 0.99 was used for the screening phase, because prostate cancer screening has little effect on short-term health status and anxiety.[17] The health states of men receiving treatment were divided into 2 months of treatment, an intermediate period (10 months of recovery from treatment), and a post-recovery period (1-10 years after treatment). Utility estimates for this post-recovery period were obtained by combining the percentage of men with side-effects from the treatment[18] with the utility estimates for those side-effects.[19] This led to a utility estimate of 0.95 for all men during the period 1-10 years after diagnosis and after receiving radical prostatectomy or radiation therapy. The loss in quality of life was calculated by

Table 8.1: Utility estimates and durations for each health state.

	Utility estimates			Source of		Source of
Health state	base	favorable	unfavorable	utility estimate	Duration	duration†
Screening attendance	0.99	1.00	0.99	[17, 20]	1 week	Assumption
Biopsy	0.90	0.94	0.87	[20]	3 weeks	Assumption
Diagnosis	0.80	0.85	0.75	based on [21]	1 month	Assumption
Radiation therapy (RT)	0.73	0.91	0.71	[19]	2 months	[19]
Radical prostatectomy (RP)	0.67	0.9	0.56	[19]	2 months	[19]
Active surveillance	0.97	1.00	0.85	[22-24]	7 years	[25]
2 months – 1 year RT	0.78	0.88	0.61	[26]	10 months	[18]
2 months – 1 year RP	0.77	0.91	0.70	[27]	10 months	[18]
Post-recovery period	0.95	1.00	0.93	[18-19]	9 years*	Assumption
Palliative therapy	0.60	0.24	0.86	[28-31]	30 months	[32]
Terminal illness	0.40	0.24	0.40	[28, 30-31]	6 months	[30-31]

*The duration of the post-recovery period used for the sensitivity analysis was the residual life-time.[33]
†Assumption refers to the authors' conclusions after discussion with experts.

multiplying the loss in utility by the duration of the health state and the number of men in that state as predicted by MISCAN.

The MISCAN model

MISCAN was used to model prostate cancer screening.[5-6] This model simulates individual life histories stochastically. The natural history of prostate cancer starts with a transition from 'no prostate cancer' to preclinical screen-detectable prostate cancer in a subset of the population. From each preclinical stage, the tumor may become screen detected, clinically diagnosed, or progress into a more advanced preclinical stage.

In the model, prostate cancers were characterized according to their clinical T-stage (T1 impalpable, T2 palpable, confined to the prostate and T3+ palpable, with extension beyond the prostatic capsule), differentiation grade (Gleason score <7, 7, or >7) and metastatic stage (locoregional or distant). The parameters for the natural history of the disease and for stage-specific test sensitivities (0.82-0.98 depending on clinical T-stage and Gleason score) were first estimated using incidence in the Dutch population during 1992-2002 (a period with limited opportunistic screening)[34] and using age and stage distributions from the Rotterdam and Gothenburg sections of the ERSPC, being the largest centers, that varied in randomization, recruitment and screening interval. In a second phase, this model was validated using screen data from all centers. The model and calibration methods and results are described in the Appendix.

Treatment assignment for locoregional cases in MISCAN was based on the age-, stage- and Gleason score-specific primary treatments (radiation therapy, radical prostatectomy and active surveillance) assigned in both arms of the ERSPC. All men with metastases and all men dying of prostate cancer were assumed to receive palliative treatment. The proportion of men receiving treatment within 7 years after having started using active surveillance was based on recent data.[35]

Survival of unscreened men diagnosed with locoregional prostate cancer was modeled using Gleason score-specific survival curves.[36] These data are from a large unscreened cohort, followed for a median period of 24 years, and the data are available by age, stage and grade. For distant disease, survival curves were based on SEER data. The effects of treatment were modeled by assuming a relative risk of dying from prostate cancer of 0.65 for radical prostatectomy[37] compared with watchful waiting. This effect was also assumed for radiation therapy.

A proportion of the screen-detected men with a locoregional cancer will be cured. In the base model, this stage-dependent cure proportion was estimated by calibrating to a prostate cancer mortality reduction of 29% after 11 years follow-up of screening at 4- year intervals for men who attended at least one screen, corresponding to the ERSPC.[2] This estimated cure proportion was used as an input to the model. Cure proportions were also estimated for hypothetical prostate cancer mortality reductions of 31% (estimated reduction adjusted for

noncompliance and contamination),[38] 35% and 39% (the intended reduction to reach of the trial in the Gothenburg center) after nine years follow-up, and of 56% after 14 years follow-up (the Gothenburg trial).[3] In the model, all screened men with prostate cancer who are cured will die from other causes at the time they would have died had they not had prostate cancer. The screened men who are not cured from prostate cancer will die at the same moment as they would have if they had not been screened. The effects of screening were calculated from 2010 until 2110, when all men will have died.

This study was designed by Heijnsdijk and de Koning. ERSPC data were gathered by each individual center and analyzed by the epidemiology committee led by Moss. Modeling was performed by Heijnsdijk, Wever, Draisma and de Koning. Quality of life data were provided by Carlsson and Korfage. The first draft was written by Heijnsdijk, with all co-authors participating in several revisions and the decision to publish the manuscript. There were no agreements concerning confidentiality of the data between the sponsors and the authors or the institutions.

RESULTS

Quality of life following treatment

Two specific studies on quality of life after prostate cancer treatment have been performed for men participating in Rotterdam and Sweden.[9, 39] Pre-operatively 1-2% of the men were incontinent and 31-40% were impotent. After 18-52 months 6-16% of the radical prostatectomy patients and 3% of the radiation therapy patients were incontinent (Table 8.2). Six to 52 months after a radical prostatectomy, 83-88% of pre-operatively potent men became impotent, compared with 42-66% of the men receiving radiation therapy. In general, screen detected men had fewer complaints both pre-operatively and postoperatively than clinically detected men (Appendix Table 8.4). This difference could be a result of aging, due to later diagnosis in the unscreened group. These ERSPC data are consistent with data from a large international cohort (Appendix Figure 8.6).

Predicted effects of annual screening at 55-69 years (base model)

The number of men experiencing each of the various health states in both the absence and presence of annual screening was modeled over the lifetime of 1,000 men (Table 8.3)..The number of life-years and QALYs gained or lost as a result of the differences between the numbers of men experiencing each health state were also calculated. The model predicted that a total of 73 life-years would be gained through the introduction of annual screening. The number of prostate cancer diagnoses was predicted to increase by screening from 112 cases to 157 cases (40% increase). The number of prostate cancer deaths was predicted to decrease from 31 to 22 (28% decrease), and the number of men receiving palliative care was predicted

Table 8.2: Frequencies of incontinence and erectile dysfunction in prostate cancer patients at two ERSPC centers at different time points.

	Treatment	Pre-operatively	6 months	12 months	18 months	52 months
Incontinence						
Regular use of pads (Gothenburg)[39]	Radical Prostatectomy n = 294	1%	NA	NA	16%	NA
Every day urinary leakage and use of 3 or more pads per day (Rotterdam)[9]	Radical Prostatectomy n = 127	2%	16%	7%	NA	6%
	Radiation Therapy n = 187	1%	1%	1%	NA	3%
*Erectile dysfunction**						
No sexual activity or impotent (Gothenburg)[39]	Radical Prostatectomy n = 294	32%	NA	NA	83%	NA
Sexually active and erectile dysfunction or sexually inactive because of erectile dysfunction (Rotterdam)[9]	Radical Prostatectomy n = 127	31%	88%	88%	NA	88%
	Radiation Therapy n = 187	40%	42%	43%	NA	66%

*The post-operative scores for erectile dysfunction represent men having normal pre-operative erectile function.
†NA denotes not available.

Table 8.3: Predicted number of men and life-years per health state comparing annual screening in men aged 55-69 years with no screening. Numbers presented are per 1,000 men aged 0-100 years, over their entire lifetime. The attendance at screening is assumed to be 80%.

		No Screening	Screening	Difference		Quality adjustment
Health state	Utility loss	(No.)	(No.)	(No.)	(Life-years)†	Life-years (range)‡
Screening attendance	- 0.01	0	8,242	8,242	158	-1.6 (-1.9, -0.3)
Biopsy	- 0.10	313	605	292	17	-1.7 (-2.2, -1.0)
Diagnosis	- 0.20	112	157	45	4	-0.7 (-0.9, -0.6)
Radiation therapy (RT)	- 0.27	43	48	5	1	-0.2 (-0.2, -0.1)
Radical prostatectomy (RP)	- 0.33	32	68	35	6	-2.0 (-2.7, -0.6)
Active surveillance	- 0.03	28	48	20	106	-3.2 (-15.8, 0)
2 months – 1 year post RT	- 0.22	43	48	5	4	-0.9 (-1.6, -0.5)
2 months – 1 year post RP	- 0.23	32	68	35	30	-6.9 (-9.1, -2.7)
Post-recovery period:	- 0.05	75	116	41		
- non-overdiagnosed					109	-5.5 (-36.4, 0)
- overdiagnosed					215	-10.8 (-30.3, 0)
Palliative therapy	- 0.40	40	26	-14	-35	14.1 (5.1, 26.9)
Terminal illness	- 0.60	31	22	-9	-4	2.6 (2.6, 3.3) +
Total life-years adjustment due to quality of life effects (sum of all health states)						-16.7 (-93.8, 24.4)
QALYs gained due to screening*						**56.0 (-20.7, 97.1)**

*Absolute number of life-years gained due to screening 72.7 years.
†The difference in the number of men in the screened and unscreened scenario is multiplied by the duration of the health states (Table 8.1).
‡The difference in life-years for each health state is multiplied by the utility loss to calculate the adjustment for quality of life.

Table 8.4: Predicted effects of various screening strategies compared with no screening.* A prostate cancer mortality reduction of 29% after 11 years using a 4 year screening interval was assumed.[40] Numbers presented are per 1,000 men aged 0-100 years, over their entire lifetime. The attendance at screening was assumed to be 80%.

	base model					
Screening age (years)	55-69	55-74	55-69	55	60	65
Screening data						
Interval (years)	1	1	4	-	-	-
Screening tests	8,242	10,577	2,250	548	584	588
Population invited†	853	891	833	685	730	735
Population screened†	845	883	777	548	584	588
Effects						
Cancers diagnosed	+ 45	+ 73	+ 29	+ 3	+ 9	+ 19
Screen detected cancers	104	150	70	8	23	42
Overdiagnosed cancers	45	72	29	2	8	19
% overdiagnosed of screen-detected	43%	48%	41%	30%	35%	45%
Negative biopsies	+ 247	+ 372	+ 166	+ 18	+ 52	+ 102
Prostate cancer deaths	- 9	- 11	- 6	- 1	- 2	- 3
Prostate cancer mortality reduction‡	37%	41%	32%	27%	29%	31%
Lead time years	1,134	1,508	750	106	262	419
Life-years (LY) gained	73	82	52	12	22	25
QALYs gained	56	56	41	12	19	17
% LY gained adjusted for quality	23%	32%	21%	6%	15%	33%
Number needed to screen	98	84	129	490	249	186
Number needed to detect	5	7	5	2	4	6

*In the unscreened scenario, there would be 112 cancers diagnosed, 201 negative biopsies and 31 prostate cancer deaths.

† The number of men invited or screened at least once.

‡The steady-state prostate cancer mortality reduction (20 years after screening) for men who have attended at least one screen.

to decrease from 40 to 26 (35% reduction). The total number of life-years gained per prostate cancer death avoided was 8.4 years (73/9). Among screened men, there was a 37% prostate cancer mortality reduction over the entire lifetime (Table 8.4).

The predicted adverse effects of screening were 247 additional negative biopsies and 41 additional men receiving radical prostatectomy or radiation therapy. The model predicted a gain of 56 QALYs (range: -21, 97), which means that (73-56)/73 = 23% of the unadjusted life-years gained would be counterbalanced by loss in quality of life. This loss was primarily attributable to the short and long-term effects of primary treatment and a longer post-recovery period with side-effects.

The number of QALYs predicted to be gained in the base model was also calculated in sensitivity analyses considering various assumptions for overdiagnosis, screening attendance, and utility estimates (Figure 8.1). A hypothetical situation without overdiagnosis was predicted to yield a gain of 79 QALYs. Screening attendance ranging from 50-100% was predicted to produce a gain of 30-60 QALYs (23% adjusted of 39 and 78 life-years gained, respectively).

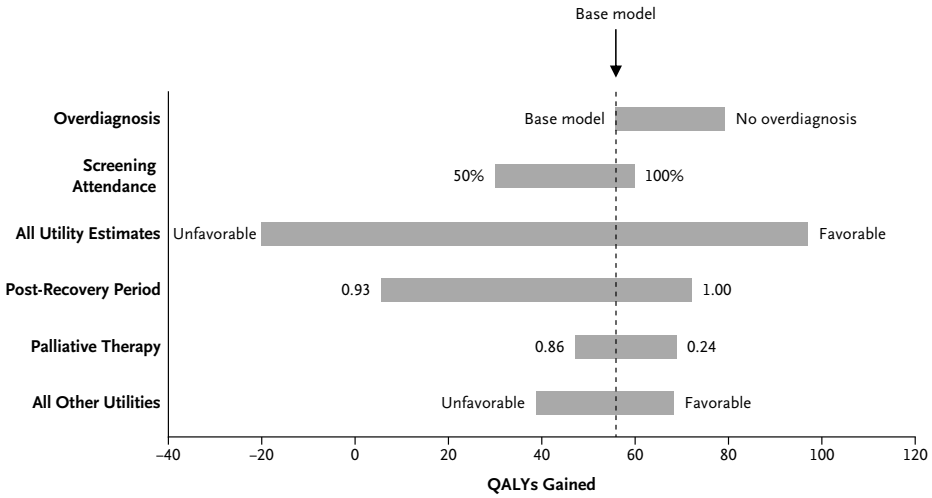


Figure 8.1: Effect of various modeling assumptions on QALYs gained in comparison with the base model (24 QALYs gained). The assumptions are: 1) no overdiagnosis, 2) screen attendance of 50% and 100%, 3) all unfavorable and favorable utility estimates, 4) utility estimate of 0.93 and 1 for the life-time post-recovery period, 5) utility estimate of 0.86 and 0.24 for palliative therapy, and 6) the utility estimates for the post-recovery period (0.95) and palliative therapy (0.6) as used in the base model combined with the unfavorable and favorable utility estimates of all other health states.

The most favorable utility estimates resulted in 97 QALYs gained, and the least favorable in 21 QALYs lost. The utility estimate for the post-recovery period had a considerable impact. If no loss in utility in this period was assumed, screening resulted in 72 QALYs gained, whereas a utility estimate of 0.93 instead of 0.95 for the remaining life-time resulted in 6 QALYs. A utility estimate of 0.95 during the first 5, 7, or 15 years after diagnosis in combination with no loss in utility after that period resulted in a gain of 66, 62 and 47 QALYs, respectively (results not shown in graph). Other utility estimates besides those for the post-recovery period and for palliative therapy had minor impact on the results.

In the base model, 104 cancers were screen detected, and 45 (43%) of these were over-diagnosed (Table 8.4). Overdiagnosed cancers are screen-detected cancers that would not have become clinically diagnosed during a person's lifetime in the absence of screening. The prostate cancer mortality reduction in a steady state (20 years after the start of screening) for men who attended at least one screening was estimated at 37%. The predicted number of men needed to screen (NNS) to prevent one prostate cancer death was 98 (845/9), and the number of men needed to detect (NND) to prevent one prostate cancer death was 5 (45/9). The predicted effects of various cure rates on the basis of various mortality reductions are described in the Appendix Table 8.5.

Predicted effects of screen strategies

Extending the screening age to 74 years resulted in an overall gain of 82 life-years and an increase in the number of prostate cancer deaths prevented from 9 to 11 (Table 8.4). However, the model predicts that only 56 QALYs (range: -47, 111) would be gained, representing a 32% reduction in unadjusted life-years. This reduction in quality of life is mainly due to the large number of overdiagnosed cases (48% of the screen detected cancers) and the 372 additional negative biopsies that would occur. On the other hand, the NNS was more favorable (84) compared with screening up to age 69.

Screening at 4-year intervals at age 55-69 years led to a gain of 52 life-years and 41 QALYs (range: -10, 69). There was a steady-state prostate cancer mortality reduction of 32% and the NNS was 129.

A single screen at age 55, 60, or 65 resulted in the detection of fewer cancers but also in less overdiagnosis. The steady-state prostate cancer mortality reduction was 27-31% and the life-years gained ranged from 12-25. The NNS for a single screen at 55, 60, or 65 years of age were 490, 249, and 186, respectively.

DISCUSSION

Weighing the balance between the benefits and harms of prostate cancer screening is essential for decision-making regarding screening at both individual and policy level. Our model predicts that there would be 9 fewer prostate cancer deaths and 73 life-years gained over the lifetime of 1,000 men using annual screening between the ages of 55–69 years. The harms caused by the introduction of such screening would be the overdiagnosis and overtreatment of 45 cases, and the loss of 1,134 prostate cancer–free life-years (lead time years). Adjusting the number of life-years gained from screening by consideration of quality of life effects showed that 56 QALYs would be gained, which is a 23% reduction from the predicted number of life-years gained.

We used a one-year screening interval in the base runs to comply with existing practice in the USA, however, the results are comparable with a 4-year interval.

The NNS (98) and NND (5) predicted in the base model are more favorable than reported in the earlier results of the ERSPC (1068 and 48, respectively).[1] The Gothenburg trial reported a NNS of 293 and a NND of 12 at 14 years follow-up.[3] Our model predicts long-term effects after a much longer period. After eleven years, the cumulative incidence of prostate cancer in the ERSPC screening group far exceeded that in the control group (9.7 versus 6.0 per 1000 person-years); however, the control group will partly catch-up because of the lead time, and therefore the absolute difference between the groups will decrease. In addition, the absolute difference in prostate cancer deaths is likely to increase over time, reducing the NNS and the NND.

A substantial part of the predicted difference between life-years and QALYs gained is caused by overdiagnosed cancers. The proportion of overdiagnosed cases (42% of the screen-detected cancers) predicted in the base model is comparable to previous studies.[6] Strategies to reduce overdiagnosis would seem to be necessary before screening can be generally advocated. Distinguishing indolent cancers from aggressive cancers, will be crucial.[42-43] More active surveillance, and deferring treatment until early signs of disease progression may also increase the QALYs gained.[44-45]

The optimal screening strategy can also depend on co-morbidity status. In our model we used general life tables for other cause mortality and therefore the distribution of co-morbidity was that of a general population. We can roughly estimate the effect of co-morbidity by adjusting the life tables. For example for men of 65 having the life-expectancy of men of 62 (low co-morbidity), annual screening from age 55-69 resulted in 93 life-years gained and 80 QALYs gained (an adjustment of 14%) and annual screening until age 75 resulted in 108 life-years gained and 86 QALYs gained (an adjustment of 20%). Therefore, screening until age 75 in men with low co-morbidity has approximately the same adjustment for quality of life as screening until age 69 in the general population.

The 23% predicted reduction in life-years gained due to quality of life effects is higher than the 8% estimated for breast cancer screening.[20] In addition to cancer deaths avoided, screening for breast cancer allows the use of less radical treatment (e.g. lumpectomy vs. mastectomy) in early detected cancers, whereas screening for prostate cancer leads to a substantial increase in treatments, especially when active surveillance strategy for indolent disease is not embraced. Also, an average of 15 life-years are gained per breast cancer death prevented while (due to older age at diagnosis and shorter life-expectancy among men) only 8.4 life-years are gained per prostate cancer death prevented.

The predicted adjustment for quality of life is due to the long-term side-effects from treatment. Both over-diagnosed and non-overdiagnosed men will live many years with adverse effects of treatment. For example, in the post-recovery period, 5 life-years were adjusted for the non-overdiagnosed men and 11 life-years for the overdiagnosed men. How these side-effects influence the long-term quality of life is not well studied. Most side-effects affecting the urinary tract and bowel will improve after some years, but significant symptoms persist in many patients up to 5 years after treatment.[18, 46-47] Although patients can adapt to these effects,[48-49] partly because they consider themselves cured from a life-threatening disease (though they could be overdiagnosed), they still report lower physical functioning 5-10 years after treatment than a control group of similar age.[46, 50-51] The results from a study of the urinary, bowel and sexual function over time after radical prostatectomy and radiation therapy, measured within the ERSPC have been compared with one of the largest studies outside the ERSPC[52] (Appendix). General patterns are similar: there is an improvement in function over time until a level slightly lower than baseline is reached (Appendix Figure 8.6). A published analysis used a decremented post-treatment utility for life-time.[33] In our base

model we used a utility estimate of 1 for the time period more than ten years after diagnosis, assuming improvement of symptoms.

One limitations of our model is that some of the utility estimates used in the present analysis are based on studies performed in the USA and these may not be representative for Europe. Also, no corrections in utility estimates were made for the detection mode (screen or clinically detected),[49] for the individual baseline quality of life level,[53] or for improvements in treatments, due to lack of detailed data. It is obvious that decreasing long-term morbidity from treatment is another important goal. However, the perceived effect of treatment on quality of life is subjective. Therefore general recommendations regarding screening do not necessarily apply to the individual.

Another limitation is that we used different datasets to develop the model. We used data from the ERSPC to estimate the parameters that are directly related with screening, or that can only be estimated from such data. For other parameters other sources were more appropriate, because of more extensive populations, more recent data or longer follow-up. We mostly used data from Rotterdam and Gothenburg, because these two large centers have different screening intervals and recruitment and therefore this variation is reflected in the model. Also, the stage distributions match well those of the entire ERSPC and they cover the entire age range. No important differences were found when PSA test sensitivities in Finland, Sweden, and the Netherlands were compared.[54]

We assumed similar effects of radiation treatment as of radical prostatectomy. No clinical trials have directly compared radical prostatectomy with radiation therapy, although some studies have shown a mortality benefit for radical prostatectomy over radiation therapy.[55-56] Assuming a relative risk of dying of 0.7 for radiation treatment would lead to an increase in the number of QALYs of a few percent.

In the NL, men have a lifetime risk of prostate cancer death of 3.5%. When screening reduces this probability with 30%, this means that 1 per 100 men would die less. This difference is too small to become statistically significant in all-cause mortality in the trial, but indeed would have an impact when screening nationwide.

In conclusion, this study quantifies how much of the benefit with the currently reported overall prostate cancer mortality reduction within ERSPC must be adjusted when the harms are taken into consideration. It is essential to await longer follow-up data from the ERSPC, as well as longer-term data on how treatment and active surveillance effects long-term quality of life before more general recommendations could be made regarding mass screening with PSA.

REFERENCES

1. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
2. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine.* Mar 15 2012;366(11):981-990.
3. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *The lancet oncology.* Aug 2010;11(8):725-732.
4. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
5. 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.* Jun 18 2003;95(12):868-878.
6. 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.* Mar 18 2009;101(6):374-383.
7. Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics.* Mar 2008;64(1):10-19.
8. Zappa M, Ciatto S, Bonardi R, Mazzotta A. Overdiagnosis of prostate carcinoma by screening: an estimate based on the results of the Florence Screening Pilot Study. *Ann Oncol.* Dec 1998;9(12):1297-1300.
9. 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.* Aug 20 2005;116(2):291-296.
10. Hugosson J, Aus G, Becker C, et al. Would prostate cancer detected by screening with prostate-specific antigen develop into clinical cancer if left undiagnosed? A comparison of two population-based studies in Sweden. *BJU Int.* Jun 2000;85(9):1078-1084.
11. Crawford ED, Abrahamsson PA. PSA-based screening for prostate cancer: how does it compare with other cancer screening tests? *European Urology.* Aug 2008;54(2):262-273.
12. Ilic D, O'Connor D, Green S, Wilt T. Screening for prostate cancer: a Cochrane systematic review. *Cancer Causes Control.* Apr 2007;18(3):279-285.
13. Roobol MJ, Schroder FH. European Randomized study of Screening for Prostate Cancer (ERSPC): rationale, structure and preliminary results. *BJU Int.* 2003;92 Suppl 2:1-122.
14. Wolters T, Roobol MJ, Steyerberg EW, et al. The effect of study arm on prostate cancer treatment in the large screening trial ERSPC. *Int J Cancer.* May 15 2010;126(10):2387-2393.
15. European Standard Population. Office for National Statistics, Newport, UK. (Accessed November 13, 2008 at http://www.statistics.gov.uk/StatBase/xsdataset.asp?vlnk=1260&More=Y)
16. The Cost-Effectiveness Analysis Registry, Center for the Evaluation of Value and Risk in Health. Boston, ICRHPS, Tufts Medical Center. (Accessed November 13, 2008 at http://www.cearegistry.org).
17. Essink-Bot ML, de Koning HJ, Nijs HG, Kirkels WJ, van der Maas PJ, Schroder FH. Short-term effects of population-based screening for prostate cancer on health-related quality of life. *J Natl Cancer Inst.* Jun 17 1998;90(12):925-931.
18. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. *N Engl J Med.* Mar 20 2008;358(12):1250-1261.
19. Stewart ST, Lenert L, Bhatnagar V, Kaplan RM. Utilities for prostate cancer health states in men aged 60 and older. *Medical care.* Apr 2005;43(4):347-355.
20. de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer.* Oct 21 1991;49(4):538-544.
21. Korfage IJ, de Koning HJ, Roobol M, Schroder FH, Essink-Bot ML. Prostate cancer diagnosis: the impact on patients' mental health. *Eur J Cancer.* Jan 2006;42(2):165-170.
22. Bennett CL, Matchar D, McCrory D, McLeod DG, Crawford ED, Hillner BE. Cost-effective models for flutamide for prostate carcinoma patients: are they helpful to policy makers? *Cancer.* May 1 1996;77(9):1854-1861.
23. Zeliadt SB, Etzioni RD, Penson DF, Thompson IM, Ramsey SD. Lifetime implications and cost-effectiveness of using finasteride to prevent prostate cancer. *The American journal of medicine.* Aug 2005;118(8):850-857.
24. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. *J Clin Oncol.* Sep 20 2011;29(27):3669-3676.
25. van den Bergh RC, Roemeling S, Roobol MJ, Roobol W, Schroder FH, Bangma CH. Prospective validation of active surveillance in prostate cancer: the PRIAS study. *European Urology.* Dec 2007;52(6):1560-1563.

26. Konski A, Sherman E, Krahn M, et al. Economic analysis of a phase III clinical trial evaluating the addition of total androgen suppression to radiation versus radiation alone for locally advanced prostate cancer (Radiation Therapy Oncology Group protocol 86-10). *Int J Radiat Oncol Biol Phys.* Nov 1 2005;63(3):788-794.
27. Calvert NW, Morgan AB, Catto JW, et al. Effectiveness and cost-effectiveness of prognostic markers in prostate cancer. *British Journal of Cancer.* Jan 13 2003;88(1):31-35.
28. Konski A, Watkins-Bruner D, Brereton H, Feigenberg S, Hanks G. Long-term hormone therapy and radiation is cost-effective for patients with locally advanced prostate carcinoma. *Cancer.* Jan 1 2006;106(1):51-57.
29. Moeremans K, Caekelbergh K, Annemans L. Cost-effectiveness analysis of bicalutamide (Casodex) for adjuvant treatment of early prostate cancer. *Value Health.* Jul-Aug 2004;7(4):472-481.
30. Penson DF, Ramsey S, Veenstra D, Clarke L, Gandhi S, Hirsch M. The cost-effectiveness of combined androgen blockade with bicalutamide and luteinizing hormone releasing hormone agonist in men with metastatic prostate cancer. *J Urol.* Aug 2005;174(2):547-552; discussion 552.
31. Ramsey S, Veenstra D, Clarke L, Gandhi S, Hirsch M, Penson D. Is combined androgen blockade with bicalutamide cost-effective compared with combined androgen blockade with flutamide? *Urology.* Oct 2005;66(4):835-839.
32. Damber JE, Aus G. Prostate cancer. *Lancet.* May 17 2008;371(9625):1710-1721.
33. 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.* Dec 1 2010;304(21):2373-2380.
34. Prostate cancer incidence and mortality rates. Comprehensive Cancer Centers, Utrecht, The Netherlands. (Accessed March 18, 2009 at http://www.ikcnet.nl).
35. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol.* Jan 1 2010;28(1):126-131.
36. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
37. 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.* Aug 20 2008;100(16):1144-1154.
38. Roobol MJ, Kerkhof M, Schroder FH, et al. Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *European Urology.* Oct 2009;56(4):584-591.
39. Carlsson S, Aus G, Bergdahl S, et al. The excess burden of side-effects from treatment in men allocated to screening for prostate cancer. The Goteborg randomised population-based prostate cancer screening trial. *Eur J Cancer.* Mar 2011;47(4):545-553.
40. Schröder FH, Hugosson J, Roobol MJ, et al. Prostate cancer mortality at 11-years of follow-up in the European Randomized study of Screening for Prostate Cancer. *N Engl J Med, in press.* 2011.
41. 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.
42. Roobol MJ, Steyerberg EW, Kranse R, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. *European Urology.* Jan 2010;57(1):79-85.
43. Steyerberg EW, Roobol MJ, Kattan MW, van der Kwast TH, de Koning HJ, Schroder FH. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol.* Jan 2007;177(1):107-112; discussion 112.
44. Stattin P, Holmberg E, Bratt O, Adolfsson J, Johansson JE, Hugosson J. Surveillance and deferred treatment for localized prostate cancer. Population based study in the National Prostate Cancer Register of Sweden. *J Urol.* Dec 2008;180(6):2423-2429; discussion 2429-2430.
45. van den Bergh RC, Roemeling S, Roobol MJ, et al. Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. *European Urology.* Jan 2009;55(1):1-8.
46. 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. *The lancet oncology.* Sep 2011;12(9):891-899.
47. 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.
48. Korfage IJ, Hak T, de Koning HJ, Essink-Bot ML. Patients' perceptions of the side-effects of prostate cancer treatment--a qualitative interview study. *Social science & medicine (1982).* Aug 2006;63(4):911-919.

49. Madalinska JB, Essink-Bot ML, de Koning HJ, Kirkels WJ, van der Maas PJ, Schroder FH. Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer. *J Clin Oncol.* Mar 15 2001;19(6):1619-1628.
50. Mols F, Korfage IJ, Vingerhoets AJ, et al. Bowel, urinary, and sexual problems among long-term prostate cancer survivors: a population-based study. *Int J Radiat Oncol Biol Phys.* Jan 1 2009;73(1):30-38.
51. Thong MS, Mols F, Kil PJ, Korfage IJ, van de Poll-Franse LV. Prostate cancer survivors who would be eligible for active surveillance but were either treated with radiotherapy or managed expectantly: comparisons on long-term quality of life and symptom burden. *BJU Int.* Mar 2010;105(5):652-658.
52. Potosky AL, Davis WW, Hoffman RM, et al. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. *J Natl Cancer Inst.* Sep 15 2004;96(18):1358-1367.
53. Chen RC, Clark JA, Talcott JA. Individualizing quality-of-life outcomes reporting: how localized prostate cancer treatments affect patients with different levels of baseline urinary, bowel, and sexual function. *J Clin Oncol.* Aug 20 2009;27(24):3916-3922.
54. Auvinen A, Raitanen J, Moss S, et al. Test sensitivity in the European prostate cancer screening trial: results from Finland, Sweden, and the Netherlands. *Cancer Epidemiol Biomarkers Prev.* Jul 2009;18(7):2000-2005.
55. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. *Cancer.* Jul 1 2011;117(13):2883-2891.
56. Cooperberg MR, Vickers AJ, Broering JM, Carroll PR. Comparative risk-adjusted mortality outcomes after primary surgery, radiotherapy, or androgen-deprivation therapy for localized prostate cancer. *Cancer.* Nov 15 2010;116(22):5226-5234.
57. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med.* Oct 27 2005;353(17):1784-1792.
58. 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.* Nov 17 2009;151(10):738-747.
59. Barton RR, Ivey JS. Nelder-Mead simplex modifications for simulation optimization. *Manage Sci.* 1996;42:954-973.
60. 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.* Aug 20 2005;116(2):291-296.

APPENDIX

The Miscan model

The Miscan model was designed at our institution and has been used extensively for the analysis and surveillance of screening programs.[1-3] In Miscan, individual life histories are first simulated in the absence of screening. A life history is defined by a sequence of states and the time spent in those states (dwelling time). The dwelling times are determined by Weibull distributions, defined by parameters for mean and shape. The states and dwelling times are generated by a semi-Markov process. From each state, a next state is generated with probabilities and dwelling times determined by the present state. Most of the transition probabilities and dwelling times are also age-dependent.

Prostate cancers were characterized according to their clinical T-stage (T1 impalpable, T2 palpable, confined to the prostate and T3+ palpable, with extension beyond the prostatic capsule), differentiation grade (Gleason score less than 7, 7, more than 7) and metastatic stage (M0 locoregional stage, M1 distant stage). In each of the states the cancer can be clinically detected or progress to the next state (Figure 1.5).

Death from other causes is generated independently using standard life tables. Screening is superimposed on the life histories in the absence of screening. Preclinical cancers can be detected by screening. Detection depends on attendance and state specific test sensitivity.

Calibration of the model

The parameters in the model for the natural history of the disease and for stage specific test sensitivities were first estimated using age and stage distribution of the cancers diagnosed in the Rotterdam and Göteborg centers of the ERSPC trial. Parameters were estimated by minimizing the difference between observed and predicted counts, measured as the sum of the chi-square quantities using the simplex method of Nelder and Mead.[4]

The following data were used: baseline incidence and stage distribution in 1991-1993 in the Netherlands, ERSPC Rotterdam trial data up to July 2004 (screen results until 2006) for both arms of the trial, baseline incidence in Sweden in 1990, and ERSPC results of Göteborg up to end of 2004. The parameters are calibrated to the baseline incidence and incidence in control arm (Appendix Table 8.1), detection rate in first and subsequent screens (Appendix Table 8.2), interval cancers (Appendix Table 8.3), clinical T-stage distribution (Appendix Figure 8.1), metastatic state (Appendix Figure 8.2) and biopsy Gleason score distribution (Appendix Figure 8.3).

In addition, the model was calibrated to the incidence of the Dutch population in 1992 – 2002 (Appendix Figure 8.4). As a validation, the predicted mortality of the Dutch population in 1992 – 2002 has been compared with the observed values (Appendix Figure 8.4). The calibrated model adequately predicted the prostate cancer incidence in the Netherlands in

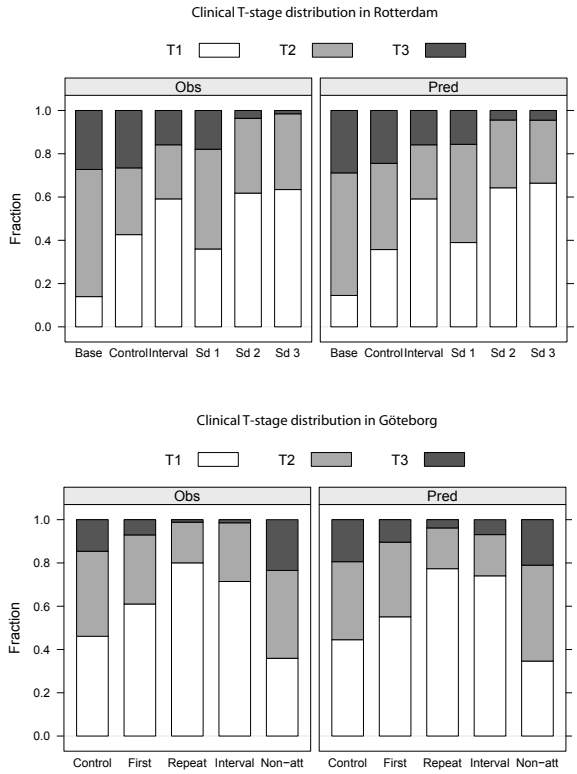
the period 1992-2002, with the exception that the prostate cancer mortality prediction was somewhat low for some age groups.

By calibrating the model to these three different datasets, we have developed a model that is applicable to multiple situations and not dependent on one of the datasets.

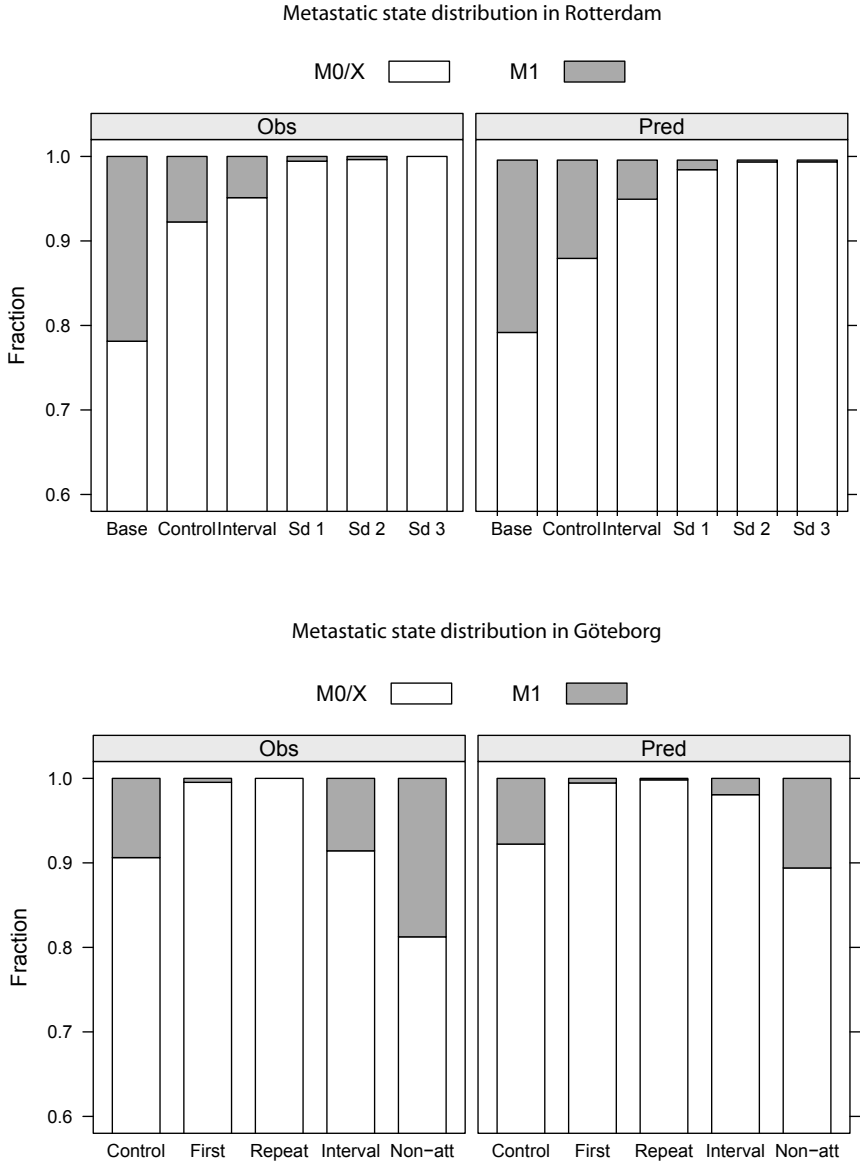
The model was validated with all screen data of all centers of the ERSPC (Appendix Figure 8.5).

Predicted effects of prostate cancer mortality reduction

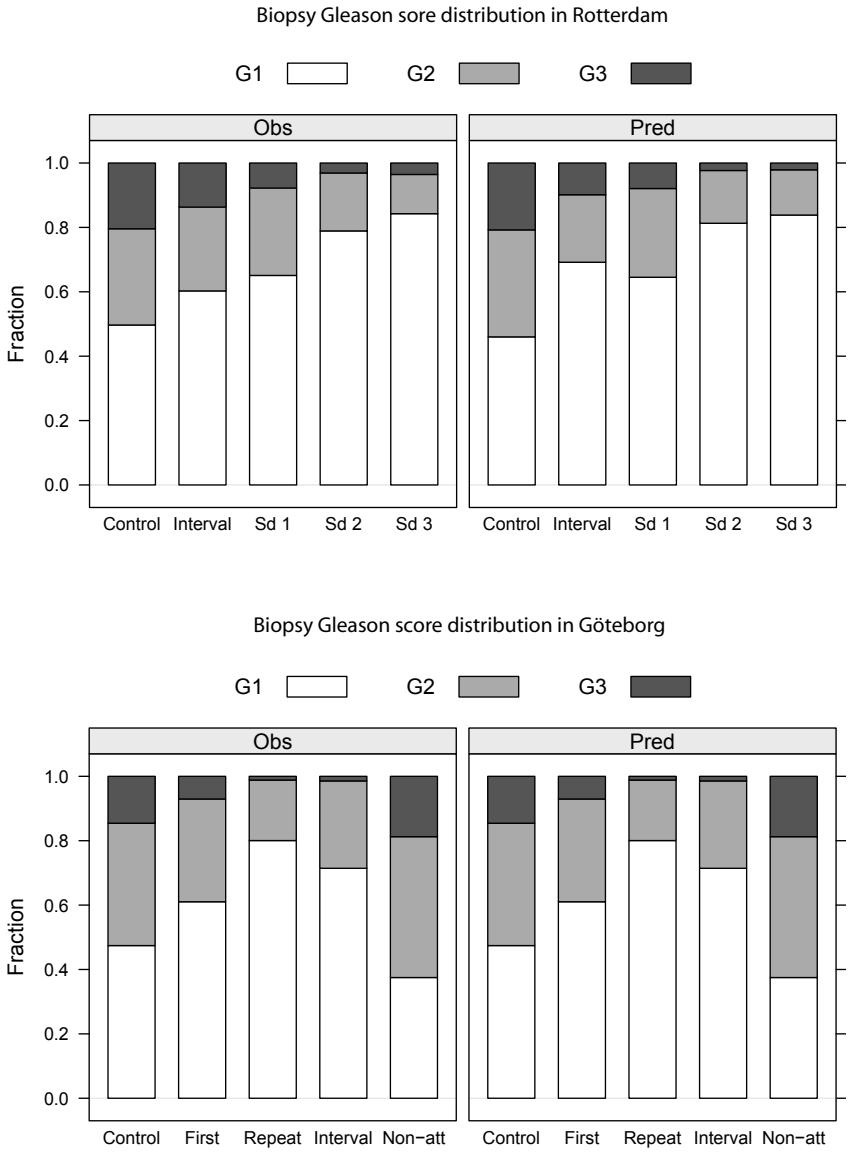
In the base model, a 29% prostate cancer mortality reduction in screened men 11 years after the start of screening was assumed based on the most recent ERSPC results. For assumed mortality reductions of 31%, 35% and 39% at nine years follow-up, and 56% after 14 years of follow-up, both the number of prostate cancer deaths avoided and the life-years and QALYs gained would increase (Appendix Table 8.5). Assuming a 56% mortality reduction, the NNS and NND would decrease to 53 and 3, respectively



Appendix Figure 8.1: Clinical T-stage distribution for the Rotterdam and Göteborg centers. For the Rotterdam center the results are presented for the baseline incidence, the control arm, the interval cancers and screening round 1-3. For the Göteborg center the results are presented for the control arm, the first and repeat screening round, the interval cancers for the attenders and the cancers detected among the non-attenders.

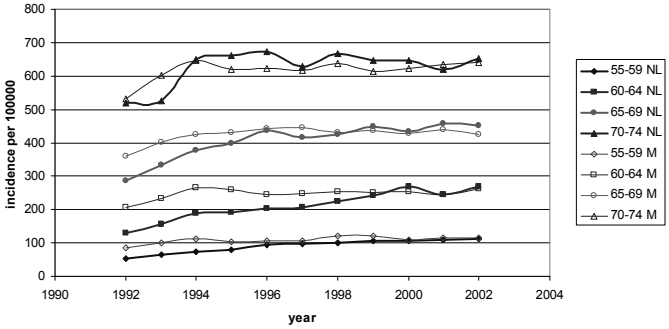


Appendix Figure 8.2: Metastatic state distribution for the Rotterdam and Göteborg centers. For the Rotterdam center the results are presented for the baseline incidence, the control arm, the interval cancers and screening round 1-3. For the Göteborg center the results are presented for the control arm, the first and repeat screening round, the interval cancers for the attenders and the cancers detected among the non-attenders.

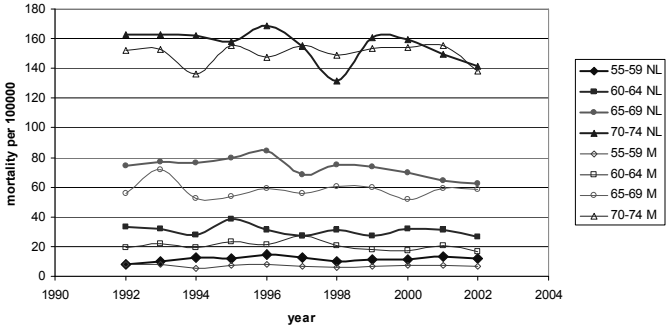


Appendix Figure 8.3: Biopsy Gleason score distribution for the Rotterdam and Göteborg centers. For the Rotterdam center the results are presented for the baseline incidence, the control arm, the interval cancers and screening round 1-3. For the Göteborg center the results are presented for the control arm, the first and repeat screening round, the interval cancers for the attenders and the cancers detected among the non-attenders.

Model predictions and observations of incidence rates in the Netherlands

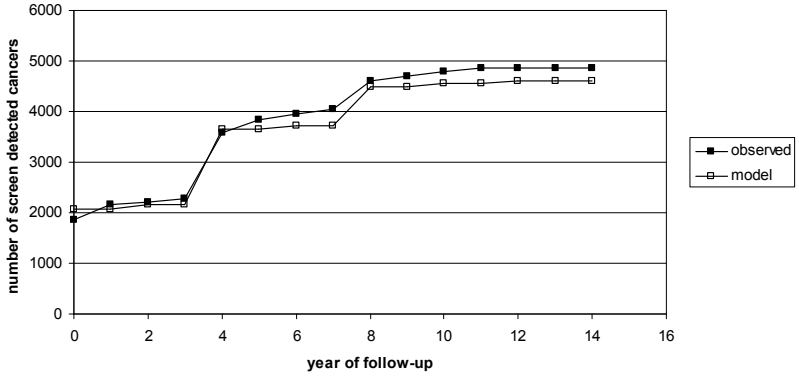


Model predictions and observations of mortality rates in the Netherlands

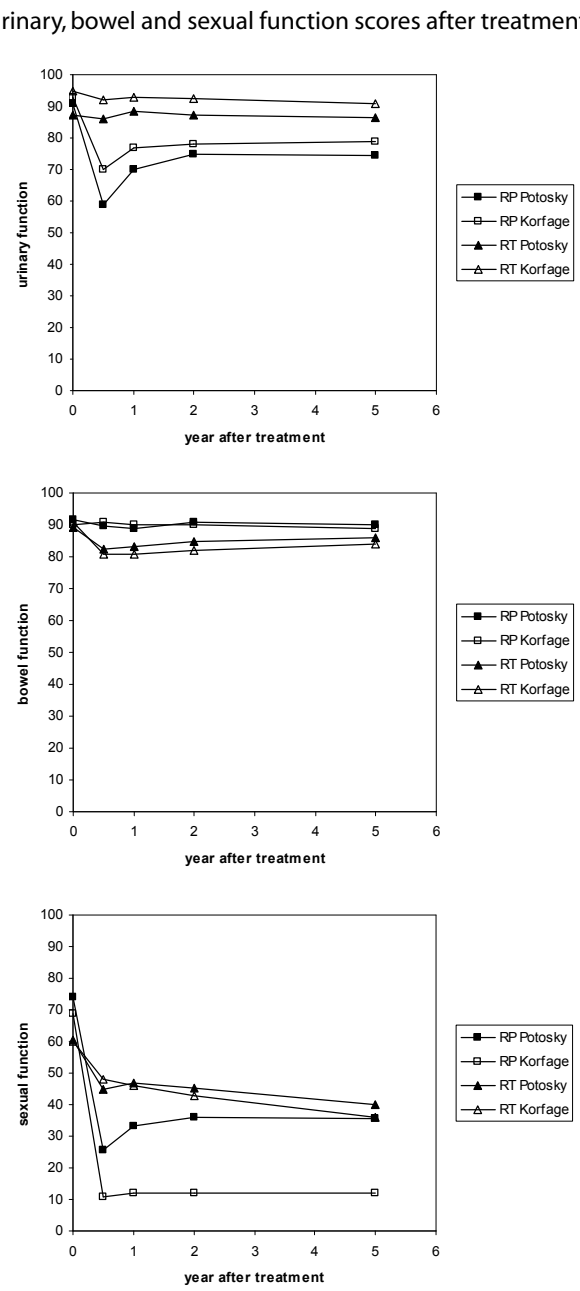


Appendix Figure 8.4: Prostate cancer incidence and mortality rates per 100,000 men in the Netherlands (bold lines) compared with the model predictions (thin lines) for the ages 55-59, 60-64, 65-69 and 70-74 in the period 1992-2002, to replicate the situation without screening.

Model predictions of the numbers of cancers detected in the screen



Appendix Figure 8.5: The observed and model-predicted number of screen detected cancers in the screen arm of all centers of the ERSPC by follow-up year.



Appendix Figure 8.6: Urinary function, bowel function and sexual function scores after radical prostatectomy and radiation therapy in years 0-5 from ERSPC Rotterdam and those published by Potosky et al (estimated values based on[5]) and Korfage et al.[6] Higher scores indicate better functioning. The scores of Potosky are obtained using the survey developed for PCOS and the scores of Korfage are obtained using the University of California, Los Angeles (UCLA), Prostate Cancer Index (PCI).

Appendix Table 8.1: The observed and model-predicted baseline incidence in Sweden and the Netherlands and the incidence in the control arm of both centers of the ERSPC. Rates per 1,000 man years.

Baseline incidence

the Netherlands			Sweden		
age group	observed	predicted	age group	observed	predicted
45-50	0.02	0.06	45-50	0.03	0.09
50-55	0.14	0.14	50-55	0.21	0.20
55-60	0.36	0.51	55-60	0.72	0.74
60-65	1.19	1.34	60-65	1.91	1.92
65-70	2.59	2.54	65-70	3.70	3.62
70-75	4.50	3.96	70-75	6.56	5.69
75-80	6.57	5.60	75-80	8.98	8.04
80-85	7.98	7.39	80-85	10.54	10.66

Incidence in control arm

Rotterdam			Göteborg		
age group	observed	predicted	age group	observed	predicted
			50-55	0.59	0.72
55-60	1.61	1.58	55-60	1.80	2.25
60-65	3.35	3.08	60-65	4.99	4.72
65-70	5.13	5.19	65-70	8.54	7.67
70-75	7.95	7.56	70-75	13.09	10.42
75-80	9.85	9.99			

Appendix Table 8.2: The observed and model-predicted detection rate by 1000 visits in the first and second screening rounds in both centers of the ERSPC.

Detection rate first round

Rotterdam			Göteborg		
age group	observed	predicted	age group	observed	predicted
			50-55	9.71	6.74
55-60	27.64	19.14	55-60	24.76	25.23
60-65	44.81	44.49	60-65	38.19	53.64
65-70	76.81	75.08	65-70	73.38	80.40
70-75	86.73	106.77			
75-80	166.67	124.89			

Detection rate second round

Rotterdam			Göteborg		
age group	observed	predicted	age group	observed	predicted
			50-55	1.31	6.10
55-60	20.77	22.05	55-60	14.43	14.76
60-65	35.14	30.17	60-65	24.39	25.18
65-70	49.74	43.74	65-70	39.38	34.30
70-75	59.30	56.69	70-75	23.62	39.24
75-80	0	68.20			

Appendix Table 8.3: The observed and model-predicted interval cancer rate by interval in years since last screen (for the Rotterdam center) and by age (for Göteborg center). The Göteborg data are divided in interval cancers rates among the attenders of at least one screen and incidence rates among the non-attenders.

Rotterdam			Göteborg				
				attenders		*non-attenders*	
interval	observed	predicted	age group	observed	predicted	observed	predicted
1	0.57	0.67	50-55	0.00	0.28	0.87	0.87
2	0.68	1.18	55-60	0.52	0.89	1.79	2.46
3	1.03	1.73	60-65	1.14	1.44	3.28	4.81
4	1.72	2.32	65-70	1.88	1.93	6.37	5.19
5	3.95	4.31	70-75	3.76	3.05	7.58	5.54

Appendix Table 8.4: Frequencies of incontinence and erectile dysfunction by detection mode in two ERSPC centers.

		pre-operative		12/18 months	
Scenario	treatment	screen-arm	control-arm	screen-arm	control-arm
Incontinence					
Regular daytime use of pads (Göteborg)	RP	0.5%	2.3%	14%	20%
Every day urinary leakage and use of 3 or more pads per day (Rotterdam)	RP	1%	3%	3%	18%
	RT	1%	0%	0%	2%
Erectile dysfunction†					
No sexual activity or impotent (Göteborg)	RP	35%	29%	79%	91%
Sexually active and erectile dysfunction or sexually inactive because of erectile dysfunction (Rotterdam)	RP	27%	38%	88%	89%
	RT	44%	37%	45%	41%

*Study arm is defined in Göteborg[7] as detected in the screen arm or control arm of the trial and in Rotterdam[6] as screen detected or clinically detected.
†For erectile dysfunction the postoperative scores are presented for the men having normal erectile function pre-operatively.

Appendix Table 8.5: Predicted effects of various assumed prostate cancer mortality reductions due to annual screening, age 55-69 years, compared with no screening. The model was based on an assumed prostate cancer mortality reduction of 29% 11 years after the start of screening,[2] 31%, 35% and 39% after 9 years and 56% after 14 years.[3] Numbers presented are per 1,000 men aged 0-100 years, over their entire lifetime. The attendance at screening was assumed to be 80%.

Model based on mortality reduction	29%	31%	35%	39%	56%
Prostate cancer deaths	-9	-10	-11	-12	-17
Prostate cancer mortality reduction*	37%	40%	46%	53%	73%
Life-years (LY) gained	73	78	90	103	136
QALYs gained	56	62	76	91	129
% LY gained adjusted for quality	23%	20%	16%	12%	5%
Number needed to screen	98	91	78	69	53
Number needed to detect	5	5	4	4	3

*The steady-state prostate cancer mortality reduction (20 years after screening) for men who have attended at least one screen.

APPENDIX REFERENCES

1. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine.* Oct 27 2005;353(17):1784-1792.
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.* Mar 18 2009;101(6):374-383.
3. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine.* Nov 17 2009;151(10):738-747.
4. Barton RR, Ivey JS. Nelder-Mead simplex modifications for simulation optimization. *Manage Sci.* 1996(42):954-973.
5. Potosky AL, Davis WW, Hoffman RM, et al. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. *J Natl Cancer Inst.* Sep 15 2004;96(18):1358-1367.
6. 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.* Aug 20 2005;116(2):291-296.
7. Carlsson S, Aus G, Bergdahl S, et al. The excess burden of side-effects from treatment in men allocated to screening for prostate cancer. The Goteborg randomised population-based prostate cancer screening trial. *Eur J Cancer.* Mar 2011;47(4):545-553.

Chapter 9

Treatment of PSA-detected prostate cancer: Benefits and harms by prognostic factors

Elisabeth M.Wever, Eveline A.M. Heijnsdijk, Gerrit Draisma, Chris H. Bangma, Monique J. Roobol, Fritz H. Schröder and Harry J. de Koning

Submitted

ABSTRACT

Background

Men with screen-detected prostate cancer have the option to treat the cancer immediately or go into an expectant management program. We present benefits and harms of immediate treatment of screen-detected local-regional prostate cancer, given the prognostic factors clinical T-stage, Gleason score and age.

Methods

A micro-simulation model, validated on available data of the ERSPC-trial, was used to predict benefits and harms of immediate curative treatment for men aged 55-74. Benefits include expected life-years gained and probability of prostate cancer death prevented. Harms include mean lead time (which is equivalent to expected loss of life-years without potential side-effects of curative treatment), and probability of over-diagnosis. Harm-benefit ratios are given for each combination of prognostic factors, comparing immediate treatment at screen-detection with treatment at the time of otherwise expected diagnosis with usual care.

Results

The ratio between mean lead time and mean life-years gained ranged from 1.8 to 31.2, and additional treatments per death prevented ranged from 0.3 to 11.6. Both ratios were lowest, so most favorable, for men aged 55-59 diagnosed with moderate-risk prostate cancer (clinical stage T3 and Gleason score <7, clinical stage T2 and Gleason score =7, clinical stage T1 and Gleason score >7, and clinical stage T2 and Gleason score >7). The ratios were high for men aged 70-74 regardless of clinical T-stage and Gleason score.

Conclusions

Men aged 55-59 years with moderate-risk prostate cancer are the best candidates for immediate curative treatment at the time of screen-detection. Immediate curative treatment is unfavorable for men aged 70-74 years regardless of clinical T-stage and Gleason score.

INTRODUCTION

The European Randomized study of Screening for Prostate Cancer (ERSPC) and the Göteborg trial (part of the ERSPC) show that a reduction in prostate mortality can be obtained with prostate-specific antigen (PSA) screening.[1-3] This mortality reduction is a result of the detection of cancers in an earlier stage, enabling treatment with curative intent. Men with a screen-detected prostate cancer have the choice between immediate active treatment and going into an expectant management program. The potential benefit of actively treating the cancer immediately after diagnosis is an increase in life-expectancy. The harm is that there is a risk of living many years with the side-effects of treatment which otherwise might have been symptom-free.[4] Alternatively, the benefits can also be expressed as the reduction of prostate cancer specific mortality and the harms as the percentage of over-diagnosis. Over-diagnosis is the number of screen-detected men with prostate cancer who in the absence of screening die from other causes before the time of clinical diagnosis as a percentage of men screen-detected with prostate cancer.

The present study quantifies the benefits and harms of immediate versus delayed active treatment per prognostic factors, clinical T-stage, Gleason scores and age, with the aim to help clinicians and patients decide whether or not to treat immediately after early detection.

Estimates for the benefits, the average life-years gained and the percentage of prostate cancer death avoided by early treatment, and the harms, the average potential life-years with no side-effects of treatment and the percentage of over-diagnosis, are obtained from a simulation model extrapolating observed rates used for parameter estimation to unobserved quantities that are of interest to patients and clinicians. These unobserved quantities can only be obtained by using models.

MATERIAL AND METHODS

ERSPC trial, Rotterdam and Göteborg section

The ERSPC trial was initiated in the early 1990s to evaluate the effect of PSA screening on prostate cancer mortality. In the Rotterdam section 42,376 men aged 55-74 years were randomized and in the Göteborg section 19,946 men aged 50-64 years. The time-interval between the screening rounds was four years in Rotterdam and two years in Göteborg. A reduction in prostate cancer mortality of 29% was observed in the overall ERSPC at a median follow-up of 11 years.[3]

Analysis

We used the MISCAN (MIcrosimulation SCreening ANalysis) prostate cancer model.[5-8] MISCAN is a micro-simulation program that simulates the progression and screening of prostate can-

Table 9.1: Model assumptions and data used in the present study.

	Assumptions	Calibration data
Other cause death	Life tables	Statistics Netherlands, 2000-2007
Disease progression before diagnosis	Semi-Markov model	National Cancer Registry data of 1991[10] Rotterdam Cancer Registry data of 1992-1993[11] Incidence in Sweden 1988-1992[12] ERSPC Rotterdam trial[5, 8] ERSPC Göteborg trial[9]
Baseline prostate-cancer-specific survival dependent on Gleason score	Poisson regression model	Connecticut Tumor Registry data[13]
Baseline prostate-cancer-specific survival dependent on clinical T-stage	Relative risk compared to baseline survival dependent on Gleason score	Population-based National Prostate Cancer Registry of Sweden[14]
Treatment effect	Relative risk compared to baseline survival	SPCG-4 study[15]
Screening effect	A proportion of men screen-detected with prostate cancer and treated early are cured	ERSPC trial[3]

cer. The model was validated on prostate cancer detection data from the ERSPC Rotterdam[5, 7-8] and the ERSPC Göteborg,[9] and on the mortality reduction data from the overall ERSPC trial.[3] A summary of the assumptions in the model and the data used for calibration are presented in Table 9.1 and are outlined in the Model appendix. For the present study we constructed a model in which individuals were screened for the first time between age 50-74 years and then subsequently every four years until age 75 years. Two situations were analyzed: one in which all men with screen-detected prostate cancer received treatment immediately at diagnosis, and the other in which they received a delayed treatment; specifically, at the time they would have been clinically diagnosed.

For the simulated populations, percentage of over-diagnosis, mean lead time, life-expectancy and percentage of prostate cancer death was calculated by the prognostic factors clinical T-stage, Gleason score and age. Using these results we calculated harm-benefit ratios: the ratio between mean lead times and mean life-years gained (= average loss of life-years without potential side-effects of curative treatment per life-year gained = M), and the ratio between the percentages of over-diagnosis and percentages of prostate cancer death avoided by treating early (= number needed to additionally treat to avoid one prostate cancer death = NNT). Note that lead time is the period by which diagnosis is advanced due to screening; therefore, mean lead time shows the average potential life-years with no side-effects of treatment in case treatment was delayed to the time of clinical diagnosis. The percentage of over-diagnosis shows the number of screen-detected men with prostate cancer who in the absence of screening die from other causes before the time of clinical diagnosis as a percentage of men screen-detected with prostate cancer.

Considering that the decision to treat immediately depends on M, we determined for which combinations of prognostic factors it is less favorable (M >9), more favorable (3 ≤ M <

9) and most favorable (M ≤ 3) to treat immediately. Note that the ranges for the three groups are chosen arbitrarily.

RESULTS

The predicted mean lead time, percentage of over-diagnosis, life-expectancy and percentage of prostate cancer deaths by the prognostic factors clinical T-stage, Gleason score and age are presented in Table 9.2 for men screen-detected with prostate cancer in the local-regional stage. To illustrate this table consider, for example, men with the following prognostic factors: diagnosed with T1G7 at age 62 years. These men have a 28.4% risk of being an over-diagnosed case. The mean lead time for such over-diagnosed men is 11.3 years, i.e. if these men decide to be treated at the time of screen-detection they live on average 11.3 years with potential side-effects of curative treatment. While, if these men had not been screened they would not have even known that they have cancer in their lifetime. Non over-diagnosed men have a mean lead time of 8.5 years, i.e. the tumor of these men is found on average 8.5 years earlier than in the absence of screening, so the tumor could be treated earlier. However, these non over-diagnosed men also have to live 8.5 years longer with the potential side-effects if they decide to be treated immediately.

The men in our example have a life-expectancy of 16.5 years and a 18.7% risk of dying from prostate cancer if the cancer is treated immediately. If the cancer is treated at the expected time of clinical diagnosis the life-expectancy is 15.3 years (1.2 years less) and the risk of dying from prostate cancer is 32.3% (13.6% more in absolute terms).

The negative effect of treating cancer at screen-detection instead of at the time of clinical diagnosis is that the patient has to live on average 9.3 years [= ((11.3*28.4) + (8.5*71.6))/100 = mean lead time for all screen-detected] longer with the potential side-effects of treatment. The positive effect is that the patient's life-expectancy is 1.2 years longer. The ratio between this negative effect and positive effect is 8.0 (due to rounding of the decimals, the ratio is not exactly 9.3/1.2). The negative effect can also be expressed as the percentage of over-diagnosed cases (28.4%) and the positive effect as the percentage of prostate cancer death avoided by treating early (13.6%). The ratio between these is the NNT (=2.1). Lower values for these ratios imply lower expected negative effects in relation to the expected positive effects and therefore plead more for immediate treatment.

Both ratios were lowest for men screen-detected with moderate risk cancer (T3G6, T2G7, T1G8 and T2G8) at age 55-59. The ratio between mean lead time and mean life-years gained was highest (31.2) for men with T3G8 at age 70-74, and NNT was highest (11.6) for men with T1G6 at age 70-74.

Figure 9.1 presents the mean life-years gained versus mean lead time for all combinations of prognostic factors. Considering that the decision to treat immediately depends on the

Table 9.2: Predicted mean lead time*, percentage of over-diagnosis†, life-expectancy‡ and percentage of pc (prostate cancer) death§ by prognostic factors: age, Gleason score and clinical T-stage. Results are for men diagnosed in the local-regional stage. The colors of the rows show for which combination of prognostic factors treatment is less favorable (dark grey, M >9), more favorable (light grey, 3 ≤ M < 9) and most favorable (white, M ≤ 3). M is mean lead time divided by mean life-years gained.

				Mean lead time (years)		Treatment at screen-detection		Treatment at symptoms presentation		M**	NNT**
Clinical T-stage	Gleason score	Age at diagnosis (years)	over-diagnosis (%)	over-diagnosed	non over-diagnosed	Life-expectancy (years)	pc death (%)	Life-expectancy (years)	pc death (%)	Mean lead time / mean life-years gained	Over-diagnosis / Δ pc death
T1	< 7	55-59	30.5	15.7	10.7	21.2	9.2	19.7	26.3	7.8	1.8
		60-64	39.9	13.4	9.5	17.6	6.5	16.7	18.7	12.1	3.3
		65-69	49.8	11.2	8.3	14.2	4.4	13.7	12.7	18.8	6.0
		70-74	60.1	8.9	6.9	10.9	2.8	10.7	8.0	30.4	11.6
	7	55-59	19.0	12.6	9.0	19.3	26.1	17.2	45.4	4.8	1.0
		60-64	28.4	11.3	8.5	16.5	18.7	15.3	32.3	8.0	2.1
		65-69	39.0	9.8	7.7	13.6	12.3	13.0	21.4	13.6	4.3
		70-74	50.1	7.9	6.7	10.7	7.4	10.4	12.9	24.3	9.2
	> 7	55-59	5.4	6.3	4.8	16.1	47.3	14.1	62.7	2.5	0.4
		60-64	9.9	6.4	4.8	14.3	37.6	13.1	49.2	4.3	0.9
		65-69	17.4	6.4	4.8	12.4	26.0	11.7	34.3	7.7	2.1
		70-74	26.6	5.5	4.4	10.0	17.6	9.6	23.1	13.1	4.8
T2	< 7	55-59	18.7	12.5	9.0	20.2	17.4	17.4	43.5	3.5	0.7
		60-64	27.9	11.2	8.5	17.0	13.0	15.4	32.2	5.7	1.5
		65-69	38.1	9.7	7.6	13.8	9.1	12.9	22.7	9.2	2.8
		70-74	49.3	7.9	6.6	10.8	6.0	10.3	14.9	14.9	5.5
	7	55-59	9.2	9.7	5.8	16.8	42.2	14.6	60.9	2.7	0.5
		60-64	15.0	8.6	5.8	14.9	32.7	13.4	47.8	4.2	1.0
		65-69	24.1	8.0	5.6	12.7	23.2	11.8	33.7	7.5	2.3
		70-74	34.5	6.8	4.9	10.2	15.1	9.8	22.0	13.5	5.0
	> 7	55-59	4.0	5.1	3.6	12.5	67.4	11.1	77.4	2.6	0.4
		60-64	6.5	4.2	3.6	11.6	57.0	10.7	65.5	3.8	0.8
		65-69	11.6	4.6	3.9	10.5	44.7	10.0	51.2	7.0	1.8

		70-74	21.1	4.7	3.8	9.0	31.3	8.7	35.9	12.8	4.5
T3	< 7	55-59	10.6	9.0	7.0	17.7	34.2	13.7	66.3	1.8	0.3
		60-64	17.7	8.5	6.9	15.3	27.5	12.8	53.4	2.8	0.7
		65-69	26.9	7.9	6.5	12.9	20.4	11.4	39.5	4.7	1.4
		70-74	38.3	6.7	5.8	10.3	13.6	9.5	26.5	8.1	3.0
	7	55-59	6.7	7.1	5.2	13.1	65.0	11.5	76.3	3.4	0.6
		60-64	11.3	6.7	5.2	12.1	55.1	11.1	64.7	5.2	1.2
		65-69	18.8	6.6	5.1	10.8	43.1	10.2	50.5	8.8	2.6
		70-74	28.8	5.8	4.7	9.1	30.3	8.8	35.5	15.0	5.5
	> 7	55-59	2.7	4.1	2.9	9.0	84.0	8.5	86.6	6.5	1.1
		60-64	4.6	3.5	2.9	8.7	76.0	8.5	78.2	10.7	2.1
		65-69	9.0	3.7	3.1	8.4	63.8	8.2	65.5	19.0	5.2
		70-74	15.5	3.9	3.2	7.6	48.9	7.5	50.4	31.2	10.5

*Lead time: interval from time of screen-detection to time of clinical diagnosis in the absence of screening (for non over-diagnosed) or to time of death (for over-diagnosed).

†Over-diagnosis: Screen-detected men with prostate cancer who in the absence of screening die from other causes before the time of clinical diagnosis as a percentage of men screen-detected with prostate cancer.

‡Life-expectancy: mean interval from screen-detection to the time of death.

§Percentage of prostate cancer death: Men screen-detected with prostate cancer who die from prostate cancer as a percentage of screen-detected men.

**Due to rounding of the decimals, ratios might be different from the ratios calculated with the estimates from the table.

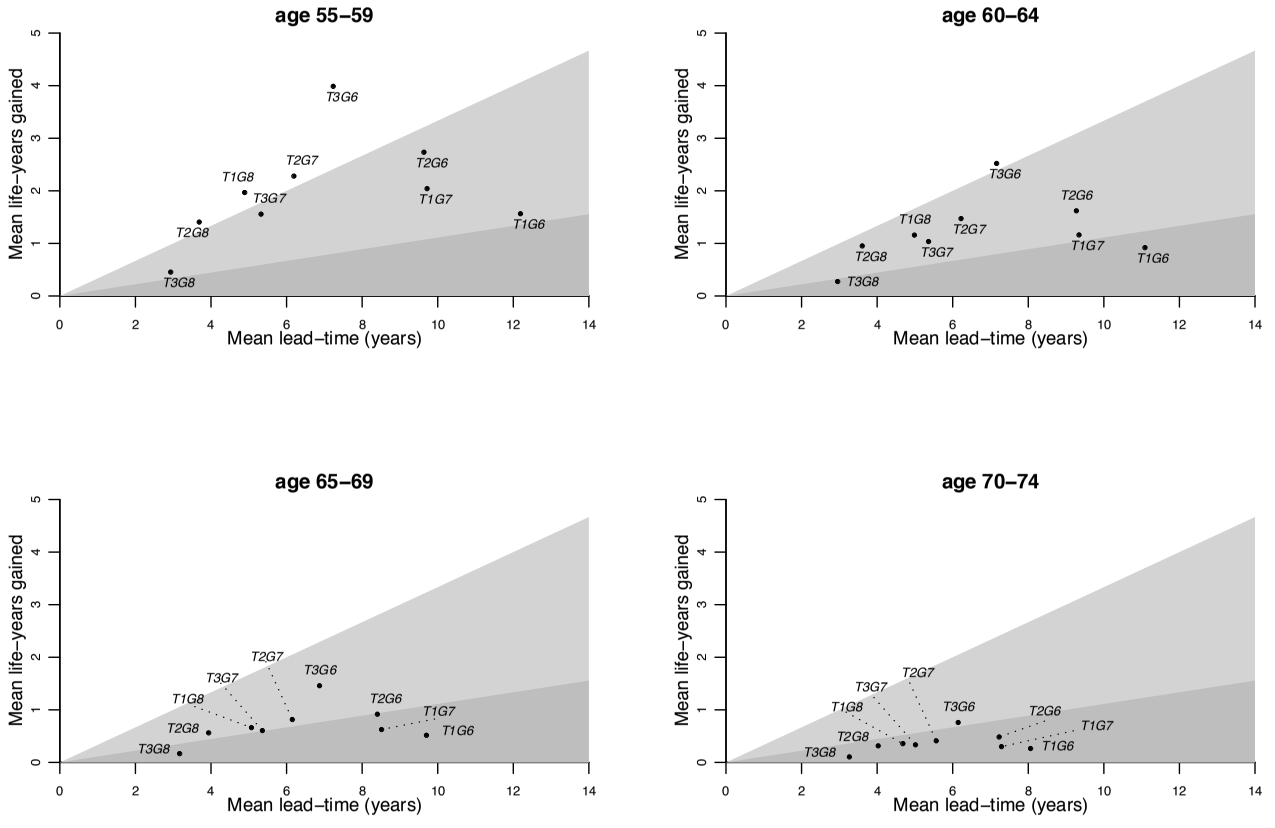


Figure 9.1: Comparison of mean life-years gained and mean lead time for different combinations of prognostic factors. For M >9 the area is dark grey (treatment less favorable), for 3 ≤ M < 9 the area is light grey (treatment more favorable) and for M ≤ 3 the area is white (treatment most favorable). M is mean lead time divided by mean life-years gained.

lead time relative to the life-years gained, this study shows for which combinations of prognostic factors it is less favorable, more favorable and most favorable to treat immediately at screen-detection. Figure 9.1 shows that, for all age groups, for patients with T3G6 immediate treatment is most favorable. While for patients with lower clinical T-stages and Gleason scores immediate treatment is less favorable because of the larger negative effects, for patients with higher clinical T-stages and Gleason scores it is less favorable because of the smaller positive effects. Therefore, for all age groups, men diagnosed with T1G6 or T3G8 are those for which immediate treatment is least favorable.

For most combinations of clinical T-stage and Gleason score patients aged 70-74 belong to the group for which the harm-benefit ratio is highest, so for which immediate treatment is not so favorable. For men aged 60-64 and 65-69 the majority belong to the group for which the harm-benefit ratio is intermediate. For men aged 55-59 for several combinations the patients belong to the group for which the harm-benefit ratio is lowest, so for which immediate treatment is most favorable.

DISCUSSION

The present study shows that, besides the potential life-years gained and the percentage of prostate cancer death avoided by early treatment, the decision to treat immediately should also depend on the lead time and the percentage of over-diagnosis. These measures depend on the prognostic factors clinical T-stage, Gleason score and age. The range of these measures given by prognostic factor is wide: Mean lead time ranges from 2.9 to 12.2 years, 2.7 to 60.1% of the screen-detected cases are over-diagnosed, mean life-years gained ranges from 0.1 to 4.0 years, and 1.5 to 32.1% of men receiving treatment at the time of screen-detection avoid prostate cancer death. These wide ranges imply that it is important to have these measures at least by these prognostic factors when deciding on whether or not to treat a patient.

The ratio between mean lead time and mean life-years gained, and the NNT, show how the harms and the benefits of immediate treatment relate to each other. These ratios are given for each combination of prognostic factors, where the ratio between mean lead time and mean life-years gained ranged from 1.8 to 31.2 and NNT ranged from 0.3 to 11.6. Both ratios were lowest for men screen-detected with moderate risk cancer, specifically for patients with T3G6 (3% of screen-detected cases at age 55-59), T2G7 (7% of screen-detected cases at age 55-59), T1G8 (1% of screen-detected cases at age 55-59) and T2G8 (2% of screen-detected cases at age 55-59). Therefore, patients with these prognostic factors are the best candidates for immediate treatment.

Men aged 70-74 with low risk cancer (T1G6) and those with high risk cancer (T3G8) had the highest values and immediate treatment is therefore least favorable for these patients. For all combinations of clinical T-stage and Gleason score, for patients aged 70-74 the harm-benefit

ratio is relatively high, which is because for these patients the probability of living long enough to enjoy the benefit of curative treatment is low. These high harm-benefit ratios for treatment also imply that the harm-benefit ratio for screening men aged 70-74 will be relatively higher than for younger men. Therefore, by only screening men who are younger the negative impact of screening might be reduced. These results imply that it is crucial to determine for which exact ages it is most favorable to screen and primarily when to stop screening.

Men with localized prostate cancer who do not treat the cancer immediately have the option to go into an expectant management program. The two expectant management approaches are active surveillance and watchful waiting. Active surveillance is an approach, where the patients will be regularly monitored with a variety of tests and if over time the disease appears to be advancing curative treatment might be performed. The aim of active surveillance is to delay or avoid treatment of prostate cancer and the potential adverse side-effects. Watchful waiting is an approach where treatment will not be given or given when symptoms appear or change. Watchful waiting is different from the active surveillance approach, in which deferred treatment has a curative intent.

Active surveillance is generally applied to men with low clinical T-stage and Gleason score.[16-19] The results presented in this study are consistent with the criteria that are in general applied. Our results indicate that patients with low clinical T-stage and Gleason score, specifically patients with T1G6, T2G6 and T1G7, might be good candidates for active surveillance; this because of the considerable harms and relatively small benefits for this group (7.2-12.2 years of lead time per 0.3-2.7 life-years gained). For patients with these prognostic factors who feel that the benefits do not outweigh the harms and therefore do not want immediate treatment, active surveillance is a good option. If these patients follow active surveillance and the disease progresses, the risk of treating an over-diagnosed cancer or a cancer which would not have been lethal during the patient's lifetime decreases and, therefore, treatment might be more favorable.

Patients with high clinical T-stage and Gleason score, who do not feel that the benefits outweigh the harms and therefore do not want immediate treatment watchful waiting might be the right approach. Also, for patients with a short life-expectancy, so for whom the probability of living long enough to enjoy the benefit of curative treatment at any time is low, watchful waiting might be the right approach.

The quantitative estimates presented are relevant for clinicians and individuals thinking about the decision whether to treat immediately at screen-detection or not. This decision should depend on how undesirable it is for the patient to live with the knowledge of having prostate cancer without treating it, and how undesirable it is for the patient to actively treat the cancer and live with the potential side-effects (e.g. sexual, urinary and bowel problems). If given the presented benefits and harms the anticipated quality of life of having been treated immediately is more than the anticipated quality of life of living under active surveillance or watchful waiting the patient should decide in favor of immediate treatment.

However, important to point out is that in this study we only compared the situation in which men screen-detected with prostate cancer receive treatment immediately at screen-detection with the one in which they receive a delayed treatment, specifically at the time they would be clinically diagnosed in the absence of screening, so with usual care. With the active surveillance and watchful waiting protocols it is probable that the cancer is treated earlier or later, which might give different values for the benefits and harms than those presented.

A limitation of the study is that the predictions are based on data from a specific population, the randomized trial in Rotterdam and Göteborg, and on assumptions underlying the model. However, using the best available data and most reasonable assumptions, we obtain the best possible estimates for the measures of interest.

Another limitation of the study is that it was not possible to consider all available predictive factors which would be available in the clinical practice. For example, our results were not presented by PSA value, MRI results or number of positive cores. Also, we did not make a distinction between Gleason score (3+4) and (4+3). These might be important prognostic factors. However, this is the first study presenting the harms of immediate treatment versus delayed treatment by the prognostic factors clinical T-stage, Gleason score and age, and it is important to have the harms and benefits by at least these prognostic factors. Also, co-morbidity is an important prognostic factor that we did not consider. However, for individuals with high co-morbidity the results of this study can be considered as regards older men.

The third limitation is that survival curves based on non-contemporary data observed in the USA was used for the ERSPC population modeled.[13] These survival data were used since it is one of the few datasets presenting survival of untreated prostate cancer as a function of age at diagnosis and Gleason score progression. The percentage of prostate cancer death might seem higher than previous reports.[13-14, 20] However, in general, the percentage of prostate cancer deaths are presented after a follow-up time of 10 or 20 years and the data presented here are lifetime data. For example, for a contemporary population, Albertsen et al. reported that after 10 years of follow-up the percentage of prostate cancer death for men with T1cG8 and age 66-75 is 13.7 to 25.7% depending on the co-morbidity.[20] In the present study, the predicted lifetime percentage of prostate cancer death for men with T1 and aged 65-74 is 27.6%; however, for 10-year follow-up this is 14.0%, which is on the low side compared with the numbers reported by Albertsen et al.[20]

In conclusion, benefits and harms of treatment immediately at PSA detection depend on the prognostic factors clinical T-stage, Gleason score and age. The range of these measures by the prognostic factors is wide. Therefore, it is important to have these measures at least by these prognostic factors when discussing about whether or not to treat. Men aged 55-59 years with moderate risk cancer (T3G6, T2G7, T1G8 and T2G8) have the lowest harm-benefit ratios and are therefore the best candidates for immediate curative treatment. Immediate curative treatment is unfavorable for men aged 70-74, regardless of the Gleason score and clinical T-stage.

REFERENCES

1. Hugosson J, Carlsson S, Aus G, et al. Mortality results from the Goteborg randomised population-based prostate-cancer screening trial. *The lancet oncology.* Aug 2010;11(8):725-732.
2. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
3. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine.* Mar 15 2012;366(11):981-990.
4. 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.* Aug 20 2005;116(2):291-296.
5. 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.* Jun 18 2003;95(12):868-878.
6. 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.* Mar 18 2009;101(6):374-383.
7. Draisma G, Postma R, Schroder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *Int J Cancer.* Nov 15 2006;119(10):2366-2371.
8. 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.* Mar 3 2010;102(5):352-355.
9. Hugosson J, Aus G, Lilja H, Lodding P, Pihl CG. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. *Cancer.* Apr 1 2004;100(7):1397-1405.
10. Visser O, Coeberg JW, Schouten LJ. Incidence of cancer in the Netherlands,1991. *Utrecht (The Netherlands): The Netherlands Cancer Registry.* 1994:13, 45.
11. Spapen SJ, Damhuis RA, Kirkels WJ. Trends in the curative treatment of localized prostate cancer after the introduction of prostate-specific antigen: data from the Rotterdam Cancer Registry. *BJU Int.* Mar 2000;85(4):474-480.
12. Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. *Cancer Incidence in Five Continents, Vol. VII* Lyon: IARC;1997.
13. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
14. Aus G, Robinson D, Rosell J, Sandblom G, Varenhorst E. Survival in prostate carcinoma--outcomes from a prospective, population-based cohort of 8887 men with up to 15 years of follow-up: results from three countries in the population-based National Prostate Cancer Registry of Sweden. *Cancer.* Mar 1 2005;103(5):943-951.
15. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.
16. Carter HB, Kettermann A, Warlick C, et al. Expectant management of prostate cancer with curative intent: an update of the Johns Hopkins experience. *J Urol.* Dec 2007;178(6):2359-2364; discussion 2364-2355.
17. Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. *Cancer.* Jun 15 2008;112(12):2664-2670.
18. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol.* Jan 1 2010;28(1):126-131.
19. van den Bergh RC, Roemeling S, Roobol MJ, et al. Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. *European Urology.* Jan 2009;55(1):1-8.
20. Albertsen PC, Moore DF, Shih W, Lin Y, Li H, Lu-Yao GL. Impact of co-morbidity on survival among men with localized prostate cancer. *J Clin Oncol.* Apr 1 2011;29(10):1335-1341.

Part 3

Discussion

Chapter 10

General Discussion

This chapter begins by addressing each of the research questions. This is followed by a general discussion focusing on the interpretation of the findings, future directions, main conclusions and recommendations for future research and practice.

10.1 ANSWERS TO RESEARCH QUESTIONS

Part 1: Modeling observed prostate cancer incidence and mortality

1. Can changes in primary treatment plausibly explain the observed decline in prostate cancer mortality by the year 2005 in the US?

Three independently developed models (MISCAN, FHCRC and UMICH) were used for this analysis. The three models project that changes in primary treatment explain 22–33% of the prostate cancer mortality decline by 2005. These contributions to mortality decline are mainly accounted for by surgery (11-14%) and radiation therapy (9-16%), both of which increased in frequency until the 1990s. The contribution made by hormonal therapies was far lower (1-3%). The results of this modeling study clearly show that changes in primary treatment are likely to have played an important role in the dramatic drop in prostate cancer mortality observed since the early 1990s. However, changes in primary treatment alone do not explain the majority of the mortality decline. The remainder of the decline is likely due to other interventions, such as PSA screening and advances in the treatment of recurrent and progressive disease.

2. Is the PSA screening performance for detecting prostate cancers in the US population the same as in the ERSPC Rotterdam?

Assuming the same natural progression of prostate cancer and sensitivity of PSA tests and consecutive biopsy in the US as in the ERSPC Rotterdam, the predicted prostate cancer incidence peak in the US was substantially higher than the prostate cancer incidence peak observed in the US population (13.3 vs. 8.1 cases per 1000 man-years). However, we could reasonably reproduce the actual observed incidence by assuming substantially lower PSA test sensitivities in the US than in ERSPC Rotterdam. For example, for nonpalpable local- or regional-stage cancers (i.e., stage T1M0), the estimates of PSA test sensitivity were 26% in the US versus 94% in ERSPC Rotterdam. These results show that PSA screening in the US did not detect as many prostate cancers as PSA screening in ERSPC Rotterdam because of the lower sensitivity of the combination of PSA testing followed by a biopsy.

3. Can we apply the commonly used assumption on the effect of screening on survival, the stage-shift assumption, to predict the observed prostate cancer mortality reduction in the ERSPC?
In stage-shift models, mortality benefits due to early detection are explained by the shift to less advanced stages, specifically to stages with more favorable stage-specific survivals. Using the stage-shift model with no additional assumption about the lead time would suggest that 90% of men with screen-detected cancers would die earlier from prostate cancer than if these cancers were clinically detected. This implausible result is obtained because part of the lead time component is neglected in the survival for screen-detected cancers. Most stage-shift models employ an additional assumption to avoid implausible results of reduced life-expectancy with screening. In this study we consider the following two stage-shift models. In the first model no death from cancer is allowed during lead time, while in the second model no death from cancer is allowed before the time of death from cancer in the absence of screening. These stage-shift models, after a follow-up of nine years, predicted reductions in prostate cancer mortality ranging from 38% to 63% for ERSPC Rotterdam, compared to a 27% reduction observed in the overall ERSPC. Therefore, using these stage-shift models to include the effect of screening considerably over-estimated the prostate cancer mortality reduction due to early detection. An alternative model, a cure model estimating stage-specific cure rates at early detection, was calibrated to correctly predict the mortality reduction (27%).

4. What is the impact of PLCO control arm contamination on perceived PSA screening efficacy?
After 10 years of complete follow-up in the PLCO trial, the average mortality rate ratios relative to an uncontaminated control arm across the three models (MISCAN, FHCRC and UMICH) were 0.68–0.80. Contamination increased the average mortality rate ratios to 0.81–0.95. We predicted that the probability that more prostate cancer deaths would have been observed in the intervention arm than in an uncontaminated control arm was 0–10%. With contamination, this probability increased to 10–30%. Contamination also decreased the power of the trial to detect a mortality difference from 40–70% (uncontaminated) to 5–25% (contaminated). Therefore, contamination substantially limited the ability of the PLCO to identify a potentially true PSA screening benefit.

Part 2: Predicting clinical outcomes of prostate cancer progression, screening and treatment

5. What are the risks of clinical progression events for PSA-detected localized prostate cases who do not receive curative treatment?
The models MISCAN, FHCRC and UMICH project that 58-86% of tumors detected by screening at age 60-64 with Gleason score 2-7 would have been diagnosed clinically (common clinical practice) in the absence of screening. For Gleason score 8-10, these risk estimates range from 84-92%. The risk that the screen-detected tumors would have metastasized before clinical

diagnosis ranges from 5-17% for tumors detected with Gleason score 2-7, and from 19%-30% for tumors detected with Gleason score 8-10. The risk that an untreated PSA-detected tumor would lead to death in the absence of any primary treatment is 20%-31% for men aged 60-64 with Gleason score 2-7, and 56%-66% for counterparts with Gleason score 8-10. These risks all decrease with age.

6. How do the lives of men who decide to be screened differ from the lives of men who decide not to be screened? And therefore, what is the anticipated loss in quality of life after a prostate cancer diagnosis and treatment that would be acceptable to decide in favor of screening?
For men who decide to be screened (annual screening from age 55 to 74), the lifetime risk of prostate cancer death decreases from 2.86% to 1.87%, while the overall life-expectancy increases by 0.08 years. At the same time, the lifetime risk of being diagnosed with prostate cancer increases from 14.10% to 21.31% and the expected time to diagnosis decreases by 1.53 years. Of the screened men, 1% enjoys the benefit of living an average of 8.08 years longer, while 17% of the screened men live an average of 9.24 years longer treated for prostate cancer. In many cases, this means an increase in life-years suffering from morbidity associated with treatment. To summarize the harms and benefits in one measure we calculated the utility break-even point. This measure implies that men who could judge that their anticipated loss in quality of life due to prostate cancer diagnosis and treatment is not more than 4.8% (range in sensitivity analysis: 0.9%-16.7%), should decide in favor of screening.

7. What are the quality of life effects of prostate cancer screening and what is the number needed to screen?
For annual screening from age 55 to 69, the predicted benefits per 1000 men of all ages were 9 fewer deaths due to prostate cancer (28% prostate cancer mortality reduction, 37% among screened men), 14 fewer men receiving palliative therapy (35% reduction), and 73 life-years gained. The harms caused by the introduction of such screening include the overdiagnosis and overtreatment of 45 men, and the loss of 1,134 prostate cancer–free life-years (lead time years). The number of QALYs gained was 56 (range -21 to 97), which means that 23% of the life-years gained would be counterbalanced by loss in quality of life. This loss was primarily attributable to the short and long-term effects of primary treatment and a longer post-recovery period with such side-effects. The predicted number of men needed to screen (NNS) to prevent one prostate cancer death was 98, and the number of men needed to detect (NND) to prevent one prostate cancer death was 5.

8. What are the benefits and harms of immediate treatment of PSA-detected local-regional prostate cancer, given the prognostic factors Gleason score, clinical T-stage and age? And for which combination of prognostic factors is immediate treatment most or least favorable?

The benefits presented are expected life-years gained and probability of prostate cancer death prevented, while the harms presented are expected loss of life-years without potential side-effects of curative treatment (which is equivalent to lead time) and probability of over-diagnosis. The range of these measures given by prognostic factor is wide: mean life-years gained ranges from 0.1 to 4.0 years; in absolute terms, 1.5% to 32.1% of men receiving treatment at the time of screen-detection avoid prostate cancer death; mean lead time ranges from 2.9 to 12.2 years; and 2.7% to 60.1% of the screen-detected cases are over-diagnosed. The ratio between mean lead time and mean life-years gained ranges from 1.8 to 31.2, and additional treatments per death prevented ranges from 0.3 to 11.6. Both harm-benefit ratios were lowest, so most favorable, for immediate treatment for men aged 55-59 with moderate risk cancer (clinical stage T3 and Gleason score <7, clinical stage T2 and Gleason score =7, clinical stage T1 and Gleason score >7, and clinical stage T2 and Gleason score >7). Therefore patients with these prognostic factors are the best candidates for immediate treatment. For older men these ratios decrease considerably since the harms decrease relatively less sharply than the benefits, implying that for older men immediate treatment is much less favorable.

10.2 INTERPRETATION OF THE FINDINGS

10.2.1 Modeling observed prostate cancer incidence and mortality

Effect of screening and treatment on observed incidence and mortality

Modeling cancer progression, screening and treatment for a population can help to understand the effects of screening and treatment on observed incidence and mortality. Without models it is often difficult to quantify the impact of screening and treatment on incidence and mortality in a population. In Chapter 2, 3 and 5 we provide important quantitative information for ongoing debates about the effects of prostate cancer screening and treatment on observed incidence and mortality in the US population and PLCO trial. In this section we will discuss these results.

Since the early 1990s we have observed a decline in prostate cancer mortality in the US: between 1991 and 2005 prostate cancer mortality declined by 42%, from 103 to 60 deaths per 100,000 men aged 50–84 years. The decrease in prostate cancer mortality coincided with several changes in the control of the disease: the widespread use of PSA tests, advances in primary treatment and changes in the recurrent and progressive disease.[1-5] Because the dissemination of PSA testing and the changes in treatment happened simultaneously, it is difficult to determine without modeling to what extent each factor contributed to the fall in prostate cancer deaths.

By modeling cancer progression, screening dissemination and treatment in the US population, we estimated that a non-trivial proportion (22–33%) of the decline in prostate cancer

mortality by year 2005 can be attributed to changes in primary treatment. However, a large proportion of the decline is left unexplained, which suggests a role for PSA screening and advances in treatment for recurrent or progressive disease.

Previously, two other CISNET prostate models[6] were used to show that early detection due to screening could account for approximately 45–70% of the decline in prostate cancer mortality under a "stage-shift" mechanism for screening benefit. The stage-shift mechanism specifies that disease shifted to an earlier stage by screening enjoys a corresponding improvement in disease-specific survival. This mechanism is a central motivator underlying many cancer screening studies. However, the study in Chapter 4 shows that using this screening-effect assumption can lead to over-estimation of the mortality reduction and that therefore less of the mortality decline might be attributed to PSA screening.

To evaluate the efficacy of prostate cancer screening, two large randomized trials were initiated in the early 1990s: the European Randomized Study of Screening for Prostate Cancer (ERSPC) in Europe and the Prostate, Lung, Colorectal, and Ovary (PLCO) trial in the United States. The ERSPC trial, at a median follow-up of eleven years, showed a significant prostate cancer mortality reduction of 29% in men actually screened,[7-8] while the PLCO trial found no statistical significant prostate cancer mortality reduction in the screen group compared with the control group.[9-10] These contradictory outcomes resulted in ongoing debates about the true benefit of PSA screening. Although the results of the two trials may seem to be in conflict, it should be noted that there were many differences in the design and implementation of the two trials. Two important differences were the diagnostic follow-up after a positive PSA test and the higher rate of screening in the control arm (contamination) in the PLCO trial.

In the PLCO trial men with a positive PSA test were referred to their personal physician for follow-up, while in the ERSPC trial they were offered biopsies at the screening centers. Therefore, the PSA screening performance in the PLCO trial is similar to that in the US population.[11] In Chapter 3 we showed that the detection rate at first PSA test was 18 cases per 1000 screened men in the US population, which is comparable to the detection rate at the first round of screening (16 per 1000 screened men) in the PLCO trial.[12] However, this study shows that if the performance of PSA screening was the same as in the ERSPC Rotterdam trial, the detection rate would have been 62 cases per 1000 screened men. To reproduce the lower detection rate in the US population, we had to assume a much lower test sensitivity of PSA screening in the US than in the ERSPC Rotterdam trial. A higher PSA cut-off level for recommending biopsy or a lower biopsy compliance rate in the US might justify this. The consequence of this lower test sensitivity is that the effects of PSA screening in the US, including in the PLCO trial, are likely to be different from those observed in the ERSPC Rotterdam trial. Therefore, although the ERSPC trial has shown that screening for prostate cancer by using PSA tests can reduce prostate cancer mortality, we cannot directly translate these mortality reductions to the US population or the PLCO trial without detailed modeling.

Some studies comparing the two trials have argued that contamination in the PLCO trial could have been sufficient to mask a statistically significant benefit or to produce a false-negative result.[13-15] In Chapter 5 we quantified the extent to which a prostate cancer mortality benefit due to screening could have been masked by contamination. Our results show that contamination substantially equalized mortality rates in the two arms, creating a non-trivial probability of more prostate cancer deaths in the intervention arm than in the control arm (increase from 0–10% (uncontaminated) to 10–30% (contaminated)), and lowering the power of the trial to detect a mortality difference between the two arms (decrease from 40–70% (uncontaminated) to 5–25% (contaminated)). Therefore, we show that contamination substantially limited the ability of the PLCO to identify a screening benefit. Before this study, the impact of contamination on the trial results had not been quantified, but it is critical for proper interpretation of the trial and synthesis of its results with those of the ERSPC and other studies.

Validation

Validation of the model has been a crucial part of the research presented in this thesis. We first modeled the ERSPC Rotterdam and Göteborg trial, and then validated the model on the data from these centers. The positive aspect about modeling and validating on randomized trial data is that the number of cases diagnosed, Gleason scores, stage distributions and number of deaths in the control arms versus those in the screen arms provide information on disease progression which is largely unobservable. The validation of the ERSPC Rotterdam and Göteborg trial provided us with estimates for the progression of prostate cancer, which we in turn used to model the incidence and mortality in the US population and PLCO trial. It is important to realize that models are only as good as the assumptions they are based on and the data they are validated on.

The results in Chapter 4 show the importance of validating models to observed randomized trial mortality outcomes. In this chapter we used stage-shift models to predict the observed prostate cancer mortality reduction in the ERSPC trial. The benefit of screening in these stage-shift models is defined by a fixed consequence of the stage distribution implied by the natural history model. These assumptions are often used without validating the model to observed trial data. When modeling the effect of screening for prostate cancer, or other cancers with long lead times and very few disease stages, the stage-shift can considerably over or under estimate the mortality reduction, and should be used with care. An alternative model, a cure model estimating stage-specific cure rates at early detection, was calibrated to correctly predict mortality reduction. The advantage of the cure model is that there is a free parameter that could be calibrated to the observed data. Another alternative model could be a stage-shift model where the correction for lead time depends on a free parameter which is estimated by calibrating to observed trial mortality outcomes.

10.2.2 Predicting clinical outcomes of prostate cancer progression, screening and treatment

Shared decision making

In health care, guidelines strongly guide decision making. In many situations the recommendations are based on a trade-off. It is important to be aware of these trade-offs and realize that patients may think differently about these trade-offs. In preference-sensitive decisions, such as the decision to be screened or treated for prostate cancer, a process of shared decision making is of great importance. These decisions are preference-sensitive because individuals assess the side-effects of prostate cancer treatment (e.g. sexual, urinary and bowel problems) very differently. Some patients find these extremely bothersome, while others claim they have no influence on their quality of life.[16-17]

In this thesis, we present clinical outcomes which provide information about the progression of prostate cancer in the absence and presence of screening and treatment. The quantitative estimates presented are relevant for clinicians and individuals thinking about the decision whether to be screened or not, and about the decision whether to be treated immediately at screen-detection or not.

The benefits and harms presented in Chapter 7 are relevant for the decision making process of screening. It is shown that screening lowers the lifetime risk of dying from prostate cancer from 2.86% to 1.87%, and therefore increases the overall life-expectancy by 0.08 years, but also that the lifetime risk of being diagnosed with prostate cancer increases from 14.10% to 21.31%, and the prostate cancer free life-years decreases by 1.53 years. The magnitude of the adverse effects depends on how undesirable it is for the individual to live with diagnosed and treated prostate cancer and the potential side-effects of treatment. In cases where an individual finds it very important to be potent or continent, and therefore thinks that the benefit is not worth the harm, he should decide not to be screened.

To summarize the harms and benefits in one measure we calculated the utility break-even point. The utility break-even point implies that men, who, based on information given, could judge that their quality of life will decrease by more than 4.8% in the event that they are diagnosed and treated for prostate cancer should probably avoid being screened. By using the utility break-even point as a threshold for screening, we would not be screening men who are unwilling to accept the side-effects from treatment, which would reduce the negative impact of screening.

For all ages, the benefits and harms to enter a screening program were quite similar. However, this does not imply that the decision to be screened should not depend on age. In Chapter 6 we only considered the common screening strategy in the US, i.e. screening annually from at least age 50 till age 75. For example, there was a large difference in the results when comparing the baseline strategy to annual screening till age 65: lifetime probability

of prostate cancer diagnosis decreased from 21.45% to 17.13%, and lifetime probability of prostate cancer death increased from 1.89% to 2.29%. (Results for the strategy screening till age 65 are only presented here.) Therefore alternative screening strategies should also be included in the decision making process. This implies that in the future, the harms and benefits of alternative screening strategies, including different intervals and age of last screening, should also be analyzed.

In Chapter 9, we also present estimates that are relevant to the shared decision making process to participate in screening and treatment. Estimates are presented for the benefits and harms of immediate versus delayed active treatment. Our results show that the benefits and harms depend on the prognostic factors clinical T-stage, Gleason score and age, and that the range of these measures by prognostic factor is wide: mean life-years gained ranged from 0.1 to 4.0 years; in absolute terms, 1.5% to 32.1% of men receiving treatment at the time of screen-detection avoided prostate cancer death; mean lead time ranged from 2.9 to 12.2 years; and 2.7% to 60.1% of the screen-detected cases were over-diagnosed; These wide ranges imply it is important to have these measures at least by these prognostic factors in discussions on whether or not to treat.

For all combinations of clinical T-stage and Gleason score for patients aged 70-74 who decide to be treated the harm-benefit ratio is relatively higher than for younger men. These high harm-benefit ratios for treatment also suggest that the harm-benefit ratio for screening men aged 70-74 will be relatively higher than for younger men. Therefore, by only screening men who are younger the negative impact of screening might be reduced. These results again imply that it is crucial to determine precisely the optimal ages to screen, and more importantly, when to stop screening.

For clinicians and individuals to actually be able to use the results of benefits and harms presented, a decision aid should be constructed. It is important that this decision aid explains the possible side-effects of diagnosis and treatment, which in turn can provide a value for the anticipated loss in quality of life for each individual. With this anticipated loss in quality of life it would then be possible to define for each individual if the expected benefits outweigh the expected harms and therefore if screening or treatment should be performed.

Lifetime risks

Randomized control trials are essential for determining the benefits and harms of a screening procedure or a treatment. However, because of the long follow-up that is needed, it is often difficult to determine the lifetime benefits and harms with only the empirical data. Nevertheless, to understand the overall benefits and harms of PSA screening and treatment it is essential to have estimates of these effects over the whole lifetime of individuals. For example, the ERSPC trial stated that,[8] after a median follow-up of eleven years, the prevention

of one prostate cancer death entailed the screening of 1,068 men (NNS: number needed to screen), and the additional diagnosis of prostate cancer in 48 men (NND: number needed to diagnose), while the estimated lifetime NNS (98) and NND (5) in Chapter 8 are much more favorable. The reported harm-benefit ratios in the ERSPC trial are quite unfavorable because of the short follow-up. After a median follow-up of eleven years, the cumulative incidence of prostate cancer in the screening group far exceeded that in the control group. However because of the lead time, it is likely that the control group will at least partially catch-up, and that the absolute difference between the groups will probably decrease. In addition, the absolute difference in prostate cancer deaths is likely to increase over time, and therefore the NNS and the NND are expected to decrease.

Quality of life

Primary treatment of prostate cancer can cause treatment related morbidity. The most prominent of which are incontinence, sexual dysfunction, and bowel problems.[18-20] These side-effects can affect an individual's quality of life tremendously.[21] Due to the high rate of over-diagnosis (43% of screen-detected cases, Chapter 8) and the prolonged course of the disease (11 lead time years for screen-detected cases, Chapter 8), it is important to analyze to what extent the harms from both overdiagnosis and treatment on quality of life can counterbalance the benefits of screening. This is important to allow effective policy decisions on population screening. In Chapter 8 we used published population average utility estimates to calculate the average quality adjusted life-years gained by screening. Our model predicts that after adjusting for quality of life, there is still a substantial benefit of screening. Specifically, it predicts that by screening those aged between 55 and 69 annually, 56 quality adjusted life-years would be gained over the lifetime of 1,000 men.

However, because of a lack of data, it is difficult to have definitive results about the benefit of screening after adjusting for quality of life. For example, due to a lack of detailed data, no corrections in utility estimates were made for the detection mode (screen or clinically detected),[22] or for improvements in treatment. Another important missing information is data on the decreasing long-term morbidity from treatment. Because of the uncertainty around utilities, the analysis was also performed using more and less favorable utilities. This sensitivity analysis predicted that between -21 to 96 quality adjusted life-years would be gained by screening. To have definitive results it is essential, therefore, to await longer-term data on the effects of treatment and active surveillance on long-term quality of life. In the Finnish centre of the ERSPC, they have been collecting quality of life data at 0, 1, 5 and 10 years from prostate cancer diagnosis from patients in both the screening and the control arm. The Prostate testing for cancer and Treatment (ProtecT) trial, a randomized trial which is ongoing to determine the effectiveness of active surveillance versus curative treatments, is also collecting quality of life data for men who receive curative treatment and for those who are in the active surveillance cohort.[23] Hopefully these studies will provide more definitive

results about how much the side-effects of treatment affect an individual's quality of life, which in turn will make it possible to have more definitive results about the quality adjusted life-years that can be gained by screening.

U.S. Preventive Services Task Force

In May 2012, the U.S. Preventive Services Task Force (USPSTF) released a recommendation against PSA-based screening for prostate cancer.[24] This recommendation was based on their opinion that there is moderate certainty that the benefits of PSA-based screening for prostate cancer do not outweigh the harms. However, several critical issues can be raised about the analysis on which this statement was based.

Firstly, in their analysis, the same value was given to both the ERSPC trial outcome (1 prostate cancer death avoided per 1000 men screened) and the PLCO trial outcome (0 prostate cancer death avoided per 1000 men screened), without considering the differences in the design and implementation of the trials. Differences in design and implementation of trials can have a big impact on outcomes. For example, in Chapter 5 we showed that opportunistic screening (contamination) substantially limited the ability of the PLCO to identify a screening benefit, and therefore the PLCO trial results published to date should not be interpreted as evidence against a benefit of PSA screening. However in the ERSPC trial, contamination was limited and therefore we can state that in this trial, the prostate cancer mortality benefit of screening versus no screening was analyzed.[25]

Secondly, the analysis of the USPSTF did not account for the fact that observed risks after some years of follow-up are very different from lifetime risks[26] (Chapter 7 and 8).

Thirdly, they did not consider the possibility of different screening protocols; limited PSA-testing might reduce the harms of screening while having only modest effect on mortality reduction.

Finally, the USPSTF states that there is a moderate certainty that the benefits of PSA-based screening for prostate cancer do not outweigh the harms. However, to make policy decisions at a population level, it is important to determine with a quantitative analysis to what extent the harms from overdiagnosis and treatment on quality of life counterbalance the benefit of screening (Chapter 8). On an individual level, it should be up to the individual's preferences and values if the potential benefit of screening outweighs the potential harms from overdiagnosis and treatment (Chapter 7).

Most other U.S. medical organizations[27-29] and the European Association of Urology[30] recommend shared decision making. However, shared decision making has not been happening in the US.[31] Many have ignored the advice of the other organizations and continue to encourage screening without providing information on the negative aspects. Therefore, despite the points of criticism on the report of the USPSTF, they deserve credit for sharpening the focus on the harms of prostate cancer screening.

Active surveillance

What is clear is that PSA screening is associated with a substantial likelihood of overdiagnosis and overtreatment, and this association places in doubt the appropriateness of aggressive therapy for every diagnosed prostate cancer patient. Active surveillance is an approach whereby patients are regularly monitored, and if over time the disease appears to be advancing, curative treatment might be performed. This approach allows one to limit prostate cancer treatment and therefore any risk of related morbidity. However, the exact benefits and harms of immediate treatment versus active surveillance remain to be quantified. The Prostate testing for cancer and Treatment (ProtecT) trial, a randomized trial, is ongoing to determine the effectiveness of active surveillance versus curative treatments.[23] Also, there are currently several ongoing prospective studies of active surveillance with specific inclusion and follow-up criteria.[32] These studies aim to determine what men choosing active surveillance can expect. Measures of interest in these studies include treatment-free survival, the number of men changing therapy, the risks of clinical progression, and cancer death. Follow-up in these studies ranges from around 2 to 7 years. Preliminary results from these cohorts show that the proportion of men moving from surveillance to active treatment ranges from 14% to 41%.[33-34] The variation in this outcome is due to different inclusion and follow-up criteria and follow-up time in the cohorts.

Several studies have already compared outcomes for men with low-risk features at diagnosis who received delayed treatment and those who received immediate therapy. For example, Dall'Era et al. showed no difference in Gleason upgrading, pathological stage and positive surgical margins between men undergoing primary radical prostatectomy and a surveillance cohort.[35] Also, Van den Bergh et al. found similar results.[36] However, it is important to note that the cohorts have been observed for too short to be able to draw definitive conclusions regarding mortality risks. Although results from the ProtecT trial and the cohort studies will take years to mature, they are expected to yield important information regarding the safety and efficacy of surveillance in a controlled setting.

While the inclusion criteria in each prospective active surveillance study are different, active surveillance is always applied to men with low risk disease, namely low clinical T-stage and Gleason score, and long life-expectancy.[37-40] The results presented in Chapter 9 are consistent with the criteria that are in general applied. Our results indicate that the best candidates for active surveillance are patients in the age group 55-59 with low clinical T-stage and Gleason score (in total 87% of screen-detected cases at age 55-59), specifically patients with T1G6 (57% of screen-detected cases at age 55-59), T2G6 (23% of screen-detected cases at age 55-59) and T1G7 (7% of screen-detected cases at age 55-59). This because of the considerable harms and relatively low benefits (9.6-12.2 lead time per 1.6-2.7 life-years gained) for this group. For patients with these prognostic factors who feel that the benefits do not outweigh the harms and therefore do not want immediate treatment immediately, active surveillance

is a good option. If these patients follow active surveillance and the disease progresses, the risk of having an over-diagnosed cancer or a cancer which would not have been lethal during the patient's lifetime decreases and, therefore, treatment might be more favorable. Patients in the age groups 60-74 with T1G6 (49% of screen-detected cases at age 60-74), T2G6 (25% of screen-detected cases at age 60-74) and T1G7 (8% of screen-detected cases at age 60-74) are also good candidates for active surveillance (in total 82% of screen-detected cases at age 60-74). However, comparing active surveillance with no curative treatment at all (watchful waiting), active surveillance is less favorable for patients aged 60-74 than patients aged 55-59 because of the lower life-expectancy.

One of the concerns of active surveillance is that patients may experience feelings of anxiety and distress while living with untreated cancer. However, anxiety in men who have chosen active surveillance or watchful waiting has not been shown to be higher than in men who elect for initial treatment.[17, 41-42] In a decision analysis, Hayes et al[43] used a state transition model to investigate the quality of life benefits and harms of active surveillance compared to initial treatments for 65 year old men with localized and low risk cancer (PSA <10 ng/ml, clinical stage ≤ T2a and Gleason score < 7). The model inputs were estimated from a systematic review, and it was additionally assumed that half of the benefit of initial curative treatment would be maintained in men undergoing active surveillance. Active surveillance was associated with the greatest quality adjusted life-expectancy (11.07 QALYs), followed by brachytherapy (10.57 QALYs), IMRT (10.51 QALYs), and radical prostatectomy (10.23 QALYs).

However, these results are highly dependent on the utility individuals place on living under active surveillance compared with having received treatment. Hayes et al. showed for several values of the utility of living after receiving treatment (assuming no side-effects) what the utility of living under active surveillance should be for active surveillance to have a higher quality adjusted life-expectancy than immediate treatment. For example, they showed that for active surveillance to have a positive effect compared to immediate treatment, the utility after receiving treatment (assuming no side-effects) should be at least 5% higher than the utility of living under active surveillance, if the baseline utility of 0.83 is assumed for the utility of living under active surveillance. In this calculation, average population utilities were assumed for the potential side-effects of treatment.

In our calculations active surveillance has a higher quality adjusted life-expectancy than immediate treatment, when the utility after treatment (including potential side-effects) is more than 5% lower than the utility of living under active surveillance. The threshold presented is for men aged 65-69 with clinical stage T1 and Gleason score < 7 and calculated with the results presented in Table 9.2. In contrast with the threshold presented by Hayes et al, when using this threshold, you should take into account that the reduction in quality of life due to the side-effects of treatment is also dependent on individual preferences.

10.2.3 Comparative modeling

Model structure is an important determinant of model predictions, but it is generally hard to assess its effect with only one model. The use of comparative modeling is therefore an important fundamental component of the CISNET approach. For the comparative modeling studies in this thesis, three independently developed models (MISCAN, FHCRC and UMICH) were used. In the studies (Chapter 2, 5 and 6), the models generated a range across the projected results. This range is a valid reflection of the uncertainty across the three models used. Despite the differences between the models, the results are usually reasonably consistent. The comparable results from different models strengthen the modeling conclusions.

Another advantage of comparative modeling is that by working collaboratively, models are improved and modeling work becomes more transparent. For example, in Chapter 6 we compared the natural history of prostate cancer in the three models, which shows important differences in the projected courses of disease development and progression. An example of conceptual differences in disease natural history of the models is manifested in projected mean years from onset of disease to clinical diagnosis. For example, the sequential stage progression formulation in the MISCAN model implies longer durations for cancers that are metastatic by the time of clinical diagnosis (mean, 13 years) than for all cancers (mean, 9 years). While in contrast, the UMICH model implicitly assumes that tumors that become metastatic by the time of clinical diagnosis tend to grow faster (mean duration to diagnosis, 4 years).

An elaborated model profile can be found on the CISNET-website for all models used in the CISNET collaboration (for all cancers).[44] These model profiles explain in detail the assumptions in the models and the data used for calibration.

10.3 FUTURE DIRECTIONS

Reconciling ERSPC and PLCO results

The ERSPC trial showed a significant prostate cancer mortality reduction of 21% in the screening group and of 29% in men actually screened.[7-8] The PLCO trial found no statistical significant prostate cancer mortality reduction in the screening group.[9-10] Although these results may seem to be in conflict, it should be noted that there were many differences in the design and implementation of the two trials. For instance, in general the ERSPC trial used a PSA threshold of 3.0 ng/ml, whereas the PLCO trial used 4.0 ng/ml for referral to biopsy. In addition, contamination (i.e., PSA screening among men in the non-screening arm) was far greater in the PLCO trial[11] than in the ERSPC trial.[45] Also, in the PLCO trial many participants had undergone prior PSA tests at baseline; this was not an issue in the ERSPC trial. Therefore, although both trials were designed to evaluate the efficacy of PSA screening, it should be recognized that the results depend on differing baseline population practices and dif-

fering compliance with differing trial protocols. The three CISNET prostate cancer models will analyze the impact of all the analyzable differences in the two trials on the efficacy of prostate cancer screening. In Chapter 5 we already made a start by analyzing the impact of contamination on the perceived PSA screening efficacy in the PLCO trial. The objective of the ongoing collaboration is to use the models to make inferences about the efficacy of PSA screening from the published trials, and to determine if the trials provide consistent evidence regarding PSA screening efficacy.

Cost-effectiveness analysis

The results of the ERSPC trial showed a significant prostate cancer mortality reduction for the PSA screening arm. The effects of screening on the lifetime probabilities of diagnosis and death, expected prostate cancer free life-years, life-expectancy and expected quality of life-years have been presented in this thesis. However before screening for prostate cancer can be implemented for the population a cost-effectiveness analysis should be performed. Therefore, future research should evaluate the quality adjusted life-years gained as well as the costs of various screening strategies and determine which strategy is most cost-effective. Alternative screening strategies, with various ages of first screening, last screening and intervals, should be considered for this analysis.

Effects of screening and treatment by co-morbidity status

Because the benefits and harms depend on co-morbidity status, it is central to prostate cancer screening and treatment decision making. Men with more co-morbidities have a higher risk of dying from a competing medical hazard than from prostate cancer.[46] Therefore, men with more co-morbidities are less likely to benefit from screening and treatment, while the harms might be substantial. For clinicians and individuals making an informed decision about screening and treatment it is therefore essential to have estimates of the benefits and harms by the prognostic factor co-morbidity-status. Hence, future research should analyze what the benefits and harms by co-morbidity-status are. The co-morbidity-status is also an important factor that should be taken into account when analyzing on a population level the optimal screening and treatment strategies.

Health disparities between black and white men

Disparities in prostate cancer outcomes among black and white men have been a source of controversy over the years, with the debate focusing on whether these disparities are due to biological differences in the disease process or to poorer access to health care among black men. While several studies have presented disparities in treatment for prostate cancer between black and white men,[47-48] these differences in care may not explain all observed disparities in prostate cancer outcomes. Even when treatment and follow-up is controlled, as in a clinical trial setting, survival differences between black and white men persist, with black

men experiencing higher mortality.[49] It is thus likely that prostate cancer may develop and progress differently in black and white men. Future research should analyze which part of the disparities is due to biological differences in the disease process and which to poorer access to health care. Subsequently, the benefits and harms for screening and treatment should be estimated for black and white men separately and it should be determined if screening and treatment policies should be different for black and white men.

Optimal active surveillance protocol

Active surveillance is gaining acceptance as an alternative initial management strategy for men with low-risk prostate cancer. It is, besides surgery and radiation, included in guidelines for the treatment of early prostate cancer.[30] However, which active surveillance protocol is optimal is still a question. In all active surveillance studies the active surveillance protocol, and the inclusion and follow-up criteria are different.[34] The outcome of these studies might provide some information about which protocol is the optimal protocol. However, these studies can not consider all possible protocols. The data from the active surveillance studies can be used to calibrate a model that estimates relevant post-diagnosis information, such as post-diagnosis PSA-growth, increase in Gleason score, metastasis progression, and prostate cancer death given the active surveillance protocol. With models like this it is then possible to compare the benefits and harms of a large variety of active surveillance protocols.

In the analysis in Chapter 9 we made a start with determining the optimal inclusion criteria. We compared strategies where men, screen-detected with prostate cancer, receive treatment immediately at screen-detection with strategies where they receive delayed treatment, specifically at the time they would be clinically diagnosed. However, the current model has its limitations. Firstly, with active surveillance protocols, it is probable that the cancer is treated earlier or later, which might give different values for the harms and benefits than those presented. Secondly, with the current model it is not possible to evaluate the optimal inclusion criteria since the relevant post-diagnosis events are not modeled. In future research, models should be constructed and validated to observed data from the active surveillance cohorts in order to determine the optimal active surveillance protocol.

10.4 MAIN CONCLUSIONS

- Changes in primary treatment explain a minority (22-33%) of the observed decline in prostate cancer mortality in the US population. The remainder of the decline is likely due to other interventions, such as PSA screening and advances in the treatment of recurrent and progressive disease.
- PSA screening in the United States did not detect as many prostate cancers as PSA screening in ERSPC Rotterdam because of the lower sensitivity of the combination of PSA testing followed by a biopsy. Therefore, the benefits and harms of PSA screening in the United States are likely to be different from those in ERSPC Rotterdam.
- Using stage-shift models to include the effect of screening on survival can considerably over-estimate the mortality reduction. Stage-shift models should be used with care, especially when modeling the effect of screening for cancers with long lead times and considering very few disease stages, such as prostate cancer.
- Control arm contamination substantially limited the ability of the PLCO trial to identify a potentially true PSA screening benefit. Therefore, the PLCO trial results published to date should not be interpreted as evidence against a benefit of PSA screening.
- Risks of disease progression among untreated PSA-detected cases can be nontrivial, particularly for younger men and men with high Gleason scores.
- Men who decide to be screened may lower their risk of dying from prostate cancer and live longer, but the associated adverse effects can be significant. The decision to be screened should take both the benefits and harms of screening into account and should depend on the personal preferences of the individual.
- Men, who, based on information given about the side-effects of treatment, might judge that their quality of life will decrease by more than 4.8% in the event that they are diagnosed and treated for prostate cancer, should probably avoid being screened. By using this threshold, we would not be screening men who are unwilling to accept the side-effects from treatment, which would reduce the negative impact of screening.
- After adjusting for quality of life, there is a benefit of screening of 56 quality adjusted life-years over the lifetime of 1,000 men using annual screening between the ages of 55–69 years. This means that 23% of the life-years gained would be counterbalanced by loss in quality of life. This loss was primarily attributable to the short and long-term effects of primary treatment and a longer post-recovery period with such side-effects.
- Benefits and harms of treatment immediately at PSA detection depend on the prognostic factors clinical T-stage, Gleason score and age. The range of these measures by the prognostic factors is wide. Therefore, it is important to have these measures at least by these prognostic factors in discussions on whether or not to treat.
- Men aged 55-59 years with T3G6, T2G7, T1G8 and T2G8 have the lowest harm-benefit ratios and are therefore the best candidates for immediate curative treatment.

10.5 RECOMMENDATIONS FOR FUTURE RESEARCH AND PRACTICE

- Changes in primary treatment alone do not explain the majority of the prostate cancer mortality decline observed since the early 1990s. Further modeling studies should investigate the extent to which screening and treatment jointly explain the mortality decline and also the role of other interventions such as advances in disease management for recurrent and metastatic disease.
- Although both the ERSPC and PLCO trial were designed to evaluate the efficacy of PSA screening, it should be recognized that the results depend on differing baseline population practices and differing compliance with differing trial protocols. The impact of the differences in the two trials on the efficacy of prostate cancer screening should be analyzed to make inferences about the efficacy of PSA screening, and to determine whether the trials provide consistent evidence regarding PSA screening efficacy.
- It should be analyzed what proportion of the disparities between black and white men in prostate cancer incidence and mortality is due to biological differences in the disease process, and what proportion is due to poorer access to health care. Consequently, benefits and harms for screening and treatment should be estimated for black and white men separately.
- Benefits and harms of screening and treatment should be quantified by co-morbidity-status. This is essential information for clinicians and individuals making an informed decision about screening and treatment.
- Because of the lack of data, it is not possible to have currently definitive results about the benefit of screening after adjusting for quality of life. Longer follow-up data from both the ERSPC and quality of life are essential before making universal recommendations regarding screening.
- Men who are deciding whether or not to be screened or treated for prostate cancer need to be fully informed on the benefits and harms of the intervention. Future research should elicit how men balance benefits and harms, and what their specific needs are with regards to informed decision making.
- Models should be constructed and validated to observed data from the active surveillance cohorts in order to determine the optimal active surveillance inclusion and follow-up criteria.
- Limited PSA-testing at population level could still appear to be a viable option. Alternative screening strategies with various ages of first screening and last screening, various intervals, and strategies with one-time screens should be considered to evaluate the quality adjusted life-years gained. The costs of these strategies should also be evaluated to determine which one is most cost-effective.

REFERENCES

1. Strope SA, Andriole GL. Prostate cancer screening: current status and future perspectives. *Nature reviews.* Sep 2010;7(9):487-493.
2. Gilliland F, Becker TM, Smith A, Key CR, Samet JM. Trends in prostate cancer incidence and mortality in New Mexico are consistent with an increase in effective screening. *Cancer Epidemiol Biomarkers Prev.* Mar 1994;3(2):105-111.
3. Hankey BF, Feuer EJ, Clegg LX, et al. Cancer surveillance series: interpreting trends in prostate cancer--part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J Natl Cancer Inst.* Jun 16 1999;91(12):1017-1024.
4. Potosky AL, Feuer EJ, Levin DL. Impact of screening on incidence and mortality of prostate cancer in the United States. *Epidemiologic reviews.* 2001;23(1):181-186.
5. Albertsen PC. The prostate cancer conundrum. *J Natl Cancer Inst.* Jul 2 2003;95(13):930-931.
6. Etzioni R, Tsodikov A, Mariotto A, et al. Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes Control.* Mar 2008;19(2):175-181.
7. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine.* Mar 26 2009;360(13):1320-1328.
8. Schroder FH, Hugosson J, Roobol MJ, et al. Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine.* Mar 15 2012;366(11):981-990.
9. Andriole GL, Crawford ED, Grubb RL, 3rd, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst.* Jan 18 2012;104(2):125-132.
10. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
11. 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.* Mar 2005;173(3):746-750; discussion 750-741.
12. Grubb RL, 3rd, Pinsky PF, Greenlee RT, et al. Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: update on findings from the initial four rounds of screening in a randomized trial. *BJU Int.* Dec 2008;102(11):1524-1530.
13. Croswell JM, Kramer BS, Crawford ED. Screening for prostate cancer with PSA testing: current status and future directions. *Oncology (Williston Park, N.Y.* May 2011;25(6):452-460, 463.
14. Schroder FH, Roobol MJ. ERSPC and PLCO prostate cancer screening studies: what are the differences? *European Urology.* Jul 2010;58(1):46-52.
15. Studer UE, Collette L. What can be concluded from the ERSPC and PLCO trial data? *Urologic oncology.* Nov-Dec 2010;28(6):668-669.
16. Korfage IJ, Hak T, de Koning HJ, Essink-Bot ML. Patients' perceptions of the side-effects of prostate cancer treatment--a qualitative interview study. *Social science & medicine (1982).* Aug 2006;63(4):911-919.
17. Steineck G, Helgesen F, Adolfsson J, et al. Quality of life after radical prostatectomy or watchful waiting. *New England Journal of Medicine.* Sep 12 2002;347(11):790-796.
18. Carlsson S, Aus G, Bergdahl S, et al. The excess burden of side-effects from treatment in men allocated to screening for prostate cancer. The Goteborg randomised population-based prostate cancer screening trial. *Eur J Cancer.* Mar 2011;47(4):545-553.
19. 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.* Aug 20 2005;116(2):291-296.
20. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. *New England Journal of Medicine.* Mar 20 2008;358(12):1250-1261.
21. Stewart ST, Lenert L, Bhatnagar V, Kaplan RM. Utilities for prostate cancer health states in men aged 60 and older. *Medical care.* Apr 2005;43(4):347-355.
22. Madalinska JB, Essink-Bot ML, de Koning HJ, Kirkels WJ, van der Maas PJ, Schröder FH. Health-related quality-of-life effects of radical prostatectomy and primary radiotherapy for screen-detected or clinically diagnosed localized prostate cancer. *J Clin Oncol.* Mar 15 2001;19(6):1619-1628.
23. Lane JA, Hamdy FC, Martin RM, Turner EL, Neal DE, Donovan JL. Latest results from the UK trials evaluating prostate cancer screening and treatment: the CAP and ProtecT studies. *Eur J Cancer.* Nov 2010;46(17):3095-3101.
24. Moyer VA. Screening for Prostate Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine.* May 21 2012.

25. Roobol MJ, Kerkhof M, Schroder FH, et al. Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *European Urology.* Oct 2009;56(4):584-591.

26. Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R. Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. *J Clin Epidemiol.* Dec 2011;64(12):1412-1417.

27. Lim LS, Sherin K. Screening for prostate cancer in U.S. men ACPM position statement on preventive practice. *American journal of preventive medicine.* Feb 2008;34(2):164-170.

28. Smith RA, Cokkinides V, Brooks D, Saslow D, Brawley OW. Cancer screening in the United States, 2010: a review of current American Cancer Society guidelines and issues in cancer screening. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):99-119.

29. Wolf AM, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA: a cancer journal for clinicians.* Mar-Apr 2010;60(2):70-98.

30. Heidenreich A, Bellmunt J, Bolla M, et al. EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and treatment of clinically localised disease. *European Urology.* Jan 2011;59(1):61-71.

31. Hoffman RM, Couper MP, Zikmund-Fisher BJ, et al. Prostate cancer screening decisions: results from the National Survey of Medical Decisions (DECISIONS study). *Arch Intern Med.* Sep 28 2009;169(17):1611-1618.

32. Dall'Era MA, Carroll PR. Outcomes and follow-up strategies for patients on active surveillance. *Current opinion in urology.* May 2009;19(3):258-262.

33. Bangma CH, Bul M, Roobol M. The Prostate cancer Research International: Active Surveillance study. *Current opinion in urology.* May 2012;22(3):216-221.

34. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. *J Clin Oncol.* Sep 20 2011;29(27):3669-3676.

35. Dall'Era MA, Cowan JE, Simko J, et al. Surgical management after active surveillance for low-risk prostate cancer: pathological outcomes compared with men undergoing immediate treatment. *BJU Int.* Apr 2011;107(8):1232-1237.

36. van den Bergh RC, Steyerberg EW, Khatami A, et al. Is delayed radical prostatectomy in men with low-risk screen-detected prostate cancer associated with a higher risk of unfavorable outcomes? *Cancer.* Mar 1 2010;116(5):1281-1290.

37. Carter HB, Kettermann A, Warlick C, et al. Expectant management of prostate cancer with curative intent: an update of the Johns Hopkins experience. *J Urol.* Dec 2007;178(6):2359-2364; discussion 2364-2355.

38. Dall'Era MA, Konety BR, Cowan JE, et al. Active surveillance for the management of prostate cancer in a contemporary cohort. *Cancer.* Jun 15 2008;112(12):2664-2670.

39. Klotz L, Zhang L, Lam A, Nam R, Mamedov A, Loblaw A. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol.* Jan 1 2010;28(1):126-131.

40. van den Bergh RC, Roemeling S, Roobol MJ, et al. Outcomes of men with screen-detected prostate cancer eligible for active surveillance who were managed expectantly. *European Urology.* Jan 2009;55(1):1-8.

41. Burnet KL, Parker C, Dearnaley D, Brewin CR, Watson M. Does active surveillance for men with localized prostate cancer carry psychological morbidity? *BJU Int.* Sep 2007;100(3):540-543.

42. van den Bergh RC, Essink-Bot ML, Roobol MJ, et al. Anxiety and distress during active surveillance for early prostate cancer. *Cancer.* Sep 1 2009;115(17):3868-3878.

43. 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.* Dec 1 2010;304(21):2373-2380.

44. National Cancer Institute, Cancer Intervention an Surveillance Modeling Network. http://cisnet.cancer.gov/. (Accessed on January 18, 2012).

45. Ciatto S, Zappa M, Villers A, Paez A, Otto SJ, Auvinen A. Contamination by opportunistic screening in the European Randomized Study of Prostate Cancer Screening. *BJU Int.* Dec 2003;92 Suppl 2:97-100.

46. Albertsen PC, Moore DF, Shih W, Lin Y, Li H, Lu-Yao GL. Impact of co-morbidity on survival among men with localized prostate cancer. *J Clin Oncol.* Apr 1 2011;29(10):1335-1341.

47. Shavers VL, Brown ML, Potosky AL, et al. Race/ethnicity and the receipt of watchful waiting for the initial management of prostate cancer. *Journal of general internal medicine.* Feb 2004;19(2):146-155.

48. Underwood W, De Monner S, Ubel P, Fagerlin A, Sanda MG, Wei JT. Racial/ethnic disparities in the treatment of localized/regional prostate cancer. *J Urol.* Apr 2004;171(4):1504-1507.

49. Albain KS, Unger JM, Crowley JJ, Coltman CA, Jr., Hershman DL. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *J. Natl. Cancer Inst.* Jul 15 2009;101(14):984-992.

Summary

Prostate cancer is the second most frequently diagnosed cancer among men worldwide. With an estimated 258,000 deaths in 2008, it is the sixth leading cause of death from cancer in men. Between 1991 and 2005, prostate cancer mortality declined by 42%, from 103 to 60 deaths per 100,000 men aged 50–84 years in the US population. The decrease in prostate cancer mortality coincided with several changes in the control of the disease: the widespread use of PSA tests, advances in primary treatment and changes in the detection and treatment of recurrent and progressive disease. Because all these changes happened simultaneously, a clear explanation for the drop in prostate cancer deaths has been elusive.

To evaluate the efficacy of prostate cancer screening, two large randomized trials were initiated in the early 1990s: the European Randomized Study of Screening for Prostate Cancer (ERSPC) in Europe and the Prostate, Lung, Colorectal, and Ovary (PLCO) trial in the United States. After a median follow-up of eleven years, the ERSPC trial, showed a significant prostate cancer mortality reduction of 29% in men actually screened, while the PLCO trial found no statistically significant prostate cancer mortality reduction in the screened group compared to the control group. These contradictory outcomes resulted in ongoing debates about the true benefit of PSA screening. Althought the results of the two trials may seem to be in conflict, it should be noted that there were many differences in the design and implementation of the two trials.

The benefit that can be achieved by screening and subsequent early treatment is a reduced risk of dying from prostate cancer. However, PSA screening is also associated with considerable unfavorable effects. Screening leads to the detection of cancers that would not have been diagnosed during the patients' lifetime if they had not been screened. Furthermore, all men with screen-detected prostate cancer who opt for curative treatment risk living many years with the side-effects of treatment which would otherwise be symptom-free years.

In this thesis, the MISCAN prostate cancer microsimulation model has been used to evaluate the effects of screening and treatment. **Part 1** covers studies that provide important quantitative information for the ongoing debates about the effects of prostate cancer on observed incidence and mortality in both the US population and the PLCO trial. In addition, it covers a study about the validation of the model. **Part 2** covers studies that present clinical outcomes which are necessary for making decisions about screening and treatment.

The aim of **Chapter 2** was to determine whether advances in primary treatment could plausibly explain the decline in prostate cancer mortality observed in the US population by the year 2005. By modeling cancer progression, screening dissemination and treatment in the US population, we estimated that 22–33% of the observed mortality decline can be attributed to changes in primary treatment. Therefore, it is shown that primary treatment explains a non-trivial fraction of the drop in disease-specific deaths. However, a large proportion of the decline is left unexplained, which suggests a role for PSA screening and advances in treatment for recurrent or progressive disease.

In **Chapter 3** we analyzed if the PSA screening performance for prostate cancer detection is the same in both the US and the ERSPC-Rotterdam. Assuming the same prostate cancer natural progression of prostate cancer and sensitivity of PSA tests and consecutive biopsy in the US as in the ERSPC Rotterdam, the predicted prostate cancer incidence peak in the US was substantially higher than the observed prostate cancer incidence peak (13.3 vs 8.1 cases per 1,000 man-years). However, the actual observed incidence could reasonably be reproduced by assuming a substantially lower PSA test sensitivity in the US than in ERSPC–Rotterdam. Therefore, we conclude that PSA screening in the US did not detect as many prostate cancers as PSA screening in ERSPC Rotterdam because of the lower sensitivity of the combination of PSA testing and consecutive biopsy. The results in this chapter imply that the effects of PSA screening in the US, including in the PLCO trial, are likely to be different from those observed in the ERSPC-Rotterdam.

The purpose of the study described in **Chapter 4** was to investigate if stage-shift assumptions, which are commonly used to include the benefit of screening on mortality in natural history models, can be used to predict the observed prostate cancer mortality reduction in the ERSPC. The stage-shift models predicted reductions in prostate cancer mortality ranging from 38% to 63% for ERSPC Rotterdam. The actual mortality reduction observed in the ERSPC was 27%. When modeling the effects of screening for prostate cancer, or other cancers which also have long lead times and very few disease stages, stage-shift models can considerably over or under-estimate the mortality reduction and thus should be used with care.

In **Chapter 5** we quantified the extent to which a prostate cancer mortality benefit due to screening could have been masked by contamination in the PLCO trial. Our results show that contamination substantially equalized mortality rates in the two arms, creating a non-trivial probability of more prostate cancer deaths in the intervention arm than in the control arm (increase from 0–10% to 10–30%), and lowering the power of the trial to detect a mortality difference between the two arms (decrease from 40–70% to 5–25%). Therefore, we show that contamination substantially limited the ability of the PLCO to identify a screening benefit.

The aim of **Chapter 6** was to determine the risks of clinical progression events for PSA-detected localized prostate cases who do not receive curative treatment. Risks of disease progression among untreated PSA-detected cases can be nontrivial, particularly for younger men and men with high Gleason scores. The models MISCAN, FHCRC and UMICH projected that 58-86% of tumors detected by screening at age 60-64 with Gleason score 2-7 would have been diagnosed clinically (common clinical practice) in the absence of screening, For Gleason score 8-10, these risk estimates range from 84-92%. The risk that an untreated PSA-detected tumor would lead to death in the absence of any primary treatment is 20%–31% for men aged 60-64 with Gleason score 2–7 and 56%–66% for counterparts with Gleason score 8–10.

The objective of **Chapter 7** was to assess the benefits and adverse effects of PSA screening for individuals who are deciding whether or not to be screened for prostate cancer. It is shown that screening increases the overall life-expectancy by 0.08 years, but also that the prostate cancer-free life-years decreases by 1.53 years. These results imply that men who, based on information provided, could judge that their quality of life would decrease by more than 4.8% if they were diagnosed and treated for prostate cancer should probably avoid being screened. By using this threshold for screening, we would not be screening men unwilling to accept the side-effects from treatment, which would reduce the negative impact of screening.

In **Chapter 8** the effects of screening on prostate cancer mortality and quality of life were quantified. Our model predicts that after adjusting for quality of life there is still a substantial benefit of screening. Specifically, it predicts that by offering annual screening to men aged between the ages of 55-69 year, 56 (range -21 to 97) quality adjusted life-years would be gained over the lifetime of 1000 men. The life-years gained would be 73 years, which means that 23% of the life-years gained would be offset by a loss in quality of life. This loss was primarily attributable to the short and long-term effects of primary treatment and a longer post-recovery period with such side-effects.

The objective of **Chapter 9** was to estimate the benefits and adverse effects of immediate treatment of PSA-detected local-regional prostate cancer compared with delayed treatment. The benefits and adverse effects are presented by the prognostic factors clinical T-stage, Gleason score and age. The range of these measures by prognostic factor is wide: mean life-years gained ranged from 0.1 to 4.0 years, and expected life-years lost without the potential side-effects of curative treatment ranged from 2.9 to 12.2 years. These wide ranges imply that it is important to have these measures, at least by these prognostic factors, when discussing with patients whether or not to treat PSA-detected local-regional prostate cancer. Men aged 55-59 years with moderate risk cancer have the lowest harm-benefit ratios and are therefore the best candidates for immediate curative treatment. For older men these ratios decrease considerably, since the harms decrease relatively less sharply than the benefits, implying that for older men immediate treatment is much less favorable.

The answers to the research questions and its implications were discussed in **Chapter 10.** Furthermore, future directions, main conclusions and recommendations for future research and practice were presented.

First we discussed the impact of prostate cancer screening and treatment on the observed incidence and mortality in both the US population and the PLCO trial. The results showed that the benefits and harms of PSA screening in the US population and the PLCO trial are likely to be different from those in the ERSPC Rotterdam trial. It is therefore important to analyze the impact of the differences in the two trials on the efficacy of prostate cancer screening to be

able to make inferences about the efficacy of PSA screening, and to determine whether the trials provide consistent evidence regarding PSA screening efficacy.

Secondly we discussed the clinical outcomes, especially the benefits and harms of PSA screening and early treatment. The ERSPC trial showed that PSA screening and subsequent treatment has the potential to decrease the rate of prostate cancer specific mortality. However, the associated negative impact can also be significant. The magnitude of the adverse effects depends largely on how undesirable it is for the individual to live with diagnosed and treated prostate cancer and its potential side-effects. Therefore, shared decision making has an important role in the decision on whether or not to be screened or treated. The clinical outcomes presented in the second part of the thesis provide important information for this decision process. The negative impact of screening might be reduced by screening men who are more willing to accept the side-effects from treatment.

It is crucial that future research investigates the most favorable age ranges to screen, paying particular attention to the optimal age at which to stop screening. Also, limited PSA-testing could appear to be a viable option to reduce the negative impact of screening. Offering an active surveillance protocol instead of immediate curative treatment to patients with low-risk cancer is also a promising strategy to reduce the negative impact of screening and consequent treatment. Although results from the active surveillance trial and cohorts will take years to mature, they are expected to yield important information regarding the safety and efficacy of surveillance in a controlled setting. Once these results are available it is important to determine the optimal active surveillance inclusion and follow-up criteria.

Samenvatting

Prostaatkanker is de tweede meest gediagnosticeerde kanker bij mannen wereldwijd. Met naar schatting 258,000 sterfgevallen in 2008 is het de zesde belangrijkste kankerdoodsoorzaak bij mannen. In de Amerikaanse bevolking daalde tussen 1991 en 2005 bij mannen in de leeftijd van 50–84 jaar de prostaatkankersterfte met 42%, van 103 naar 60 doden per 100,000 mannen. De afname van prostaatkankersterfte viel samen met diverse veranderingen in de controle van de ziekte: het uitgebreid gebruik van PSA testen, de vooruitgang in primaire behandelingen en de veranderingen in de opsporing en behandeling van terugkerende en progressieve kanker. Omdat al deze veranderingen gelijktijdig gebeurd zijn, is een duidelijke verklaring voor de daling in de prostaatkankersterfte niet vanzelfsprekend.

Om de doeltreffendheid van vroege opsporing van prostaatkanker te evalueren zijn aan het begin van de jaren 90 twee grote gerandomiseerde trials gestart: De "European Randomized Study of Screening for Prostate Cancer" (ERSPC) in Europa en de "Prostate, Lung, Colorectal, and Ovary" (PLCO) trial in de Verenigde Staten. De ERSPC trial toonde een aanzienlijke prostaatkankersterfte vermindering van 29% aan bij de gescreende mannen na een mediaan follow-up periode van 11 jaar. Echter vond men in de PLCO trial geen statistisch significante vermindering in prostaatkankersterfte in de gescreende groep in vergelijking met de controle groep. Deze tegenstrijdige resultaten resulteerden in aanhoudende discussies over het echte voordeel van screenen met de PSA test. Hoewel de resultaten van de twee trials in conflict lijken te zijn, moet er opgemerkt worden dat er veel verschillen waren in het ontwerp en uitvoering van de twee trials.

Het voordeel dat kan worden bereikt door het vroeg opsporen en vroeg behandelen van prostaatkanker is een verlaagd risico om dood te gaan aan prostaatkanker. Echter, zijn er ook aanzienlijke ongunstige effecten geassocieerd met PSA screening. Screening leidt namelijk ook tot de ontdekking van kankers die niet zouden worden ontdekt gedurende het hele leven van de persoon als hij niet gescreend zou zijn (over-diagnose). Bovendien leven alle mannen die ontdekt zijn door screening en die voor curatieve behandeling kiezen vele jaren met bijwerkingen van de behandeling. Deze jaren zouden anders symptoom vrije jaren kunnen zijn geweest.

In dit proefschrift is het MISCAN prostaatkanker microsimulatie model gebruikt om de effecten van vroege opsporing en behandeling te evalueren. **Deel 1** heeft betrekking op studies die belangrijke kwantitatieve informatie leveren voor lopende discussies over de impact van screening en behandeling op de waargenomen prostaatkanker incidentie en mortaliteit in de Amerikaanse bevolking en PLCO trial. Bovendien omvat het een studie over de validatie van het model. **Deel 2** heeft betrekking op studies die klinische resultaten presenteren die nodig zijn bij het nemen van beslissingen over screening en behandeling.

Het doel van **Hoofdstuk 2** was om te bepalen of de waargenomen daling in prostaatkankersterfte in de VS te verklaren is door de vooruitgang in primaire behandeling. Door het modelleren van de progressie van kanker, de verspreiding van screening en de behandeling

van prostaatkanker in de VS, hebben we geschat dat 22-33% van de prostaatkankersterfte kan worden toegeschreven aan veranderingen in primaire behandelingen. Hieruit volgt dat een groot deel van de daling onverklaard blijft. Dit suggereert een rol voor PSA screening en vooruitgang in de behandelingen van terugkerende en progressieve ziekte.

In **Hoofdstuk 3** hebben we geanalyseerd of de prestaties van PSA screening met betrekking tot de opsporing van prostaatkanker in de VS gelijk is aan die in de ERSPC trial. Als we aannemen dat de natuurlijke progressie van kanker en de sensitiviteit van de PSA test met opeenvolgende biopsie in de VS hetzelfde zijn als in de ERSPC Rotterdam, schatten we een incidentie piek in de VS dat aanzienlijk hoger is dan de waargenomen prostaatkanker incidentie piek (13.3 vs 8.1 gevallen per 1000 manjaren). De waargenomen incidentie kan echter redelijk worden gereproduceerd door een aanzienlijk lagere PSA test sensitiviteit aan te nemen in de VS. Daarom kunnen we concluderen dat de detectiecijfer van PSA screening in de VS lager is dan in ERSPC Rotterdam als gevolg van een lagere sensitiviteit van de PSA test en opeenvolgende biopsie. De resultaten in dit hoofdstuk impliceren dat de effecten van PSA screening in de VS, inclusief in de PLCO trial, waarschijnlijk anders zullen zijn dan die waargenomen in de ERSPC trial.

Het doel van **Hoofdstuk 4** was om te onderzoeken of aannames die vaak worden gebruikt om het effect van screening op sterfte te modelleren, "stage-shift" aannames, kunnen worden gebruikt om de waargenomen reductie in prostaatkankersterfte in de ERSPC te modelleren. De stage-shift modellen voorspelden een reductie in prostaatkankersterfte variërend van 38% tot 63% in vergelijking met een 27% waargenomen reductie in de ERSPC trial. Bij het modelleren van het effect van screening voor kankers met lange lead times, die bovendien heel weinig stadia van de ziekte onderscheiden, zoals prostaatkanker, kunnen stage-shift modellen aanzienlijk de mortaliteit reductie over of onder schatten en moeten daarom zorgvuldig worden gebruikt.

In **Hoofdstuk 5** hebben we gekwantificeerd in welke mate contaminatie in de controle groep van de PLCO trial een voordeel in prostaatkankersterfe als gevolg van screening zou kunnen hebben gemaskeerd. Onze resultaten laten zien dat door contaminatie de geschatte prostaatkankersterfte in de controle groep veel dichter bij die van de gescreende groep komt. We hebben ook geschat dat de kans 0-10% is dat er meer prostaatkankersterfte zou zijn geweest in de gescreende groep dan in een niet-gecontamineerde controle groep. Door contaminatie in de controle groep is deze kans verhoogd tot 10-30%. Ook laten we zien dat contaminatie in de controle groep de power van de trial om een verschil in prostaatkankersterfte te identificeren van 40–70% (niet-gecontamineerd) naar 5–25% (gecontamineerd) heeft laten afnemen. Met deze resultaten tonen we aan dat contaminatie het vermogen van de PLCO trial om een voordeel van screening te identificeren behoorlijk beperkt heeft.

Het doel van **Hoofdstuk 6** was om het risico op klinische progressie te bepalen voor gelokaliseerde prostaatkanker gevallen die door middel van PSA screening zijn gedetecteerd en die niet curatief behandeld zijn. De modellen MISCAN, FHCRC and UMICH hebben geprojecteerd dat 58-85% van de kankers, gedetecteerd door screening in de leeftijd van 60-64 jaar, met een Gleason score van 2-7 op den duur klinisch (gangbare klinische praktijk) gediagnosticeerd zouden worden in het geval dat er geen screening zou zijn. Voor een Gleason score van 8-10 varieert deze risicoschatting van 84%-92%. Het risico dat een onbehandelde PSA gedetecteerde tumor zou leiden tot het overlijden aan prostaatkanker is 20-31% voor mannen in de leeftijd van 60-64 jaar met een Gleason score van 2-7, en 56-66% voor mannen met een Gleason score van 8–10.

De doelstelling van **Hoofdstuk 7** was om de positieve en negatieve effecten van PSA screening te bepalen voor individuen die aan het beslissen zijn of ze wel of niet willen worden gescreend. De resultaten laten zien dat door het vroeg opsporen en vroeg behandelen van prostaatkanker de levensverwachting met 0.08 jaren toeneemt, maar ook dat de prostaatkanker-vrije levensjaren met gemiddeld 1.53 jaren afnemen. Deze resultaten impliceren dat mannen zich niet moeten laten screenen als ze kunnen beoordelen dat hun kwaliteit van leven met meer dan 4.8% zal afnemen wanneer ze gediagnosticeerd en vervolgens behandeld worden voor prostaatkanker. Door deze drempelwaarde te gebruiken, worden mannen die niet bereid zijn om de bijwerkingen van de behandeling te accepteren niet gescreend en zou de negatieve impact van screenen kunnen worden gereduceerd.

In **Hoofdstuk 8** presenteren we de effecten van screening op prostaatkankersterfte en kwaliteit van leven. Ons model voorspelt dat na rekening te houden met de kwaliteit van leven er nog steeds een aanzienlijk voordeel van screening is. Namelijk, dat bij het screenen van 1,000 mannen in de leeftijd van 55-69 jaar er na het corrigeren voor kwaliteit van leven 56 (range -21 tot 97) levensjaren (“quality adjusted life-years”) te winnen zijn. De winst in levensjaren was 73 jaren, wat betekent dat het tegeneffect door de vermindering van de kwaliteit van leven door prostaatkanker screening en behandeling 23% van de gewonnen levensjaren is. Dit verlies was voornamelijk toe te schrijven aan de korte en lange termijn effecten van primaire behandeling en een lange post-recovery periode met dergelijke bijwerkingen.

De doelstelling van **Hoofdstuk 9** was om de positieve en negatieve effecten van onmiddellijke behandeling versus uitgestelde behandeling te schatten voor mannen met PSA gedetecteerde lokale-regionale prostaatkanker. De positieve en negatieve effecten zijn gepresenteerd per prognostische factor klinische T stadium, Gleason score en leeftijd. Het bereik van de effecten gegeven de prognostische factoren is breed: de gemiddelde gewonnen levensjaren varieerde van 0.1 tot 4.0 jaar en het gemiddelde verlies van levensjaren zonder potentiële bijwerkingen van curatieve behandeling varieerde van 2.9 to 12.2 jaar. Deze brede

bereiken geven aan dat het belangrijk is om deze waardes in ieder geval per prognostische factor klinische T stadium, Gleason score en leeftijd te hebben. Mannen in de leeftijd van 55-59 jaar met een matig risico kanker hebben de laagste schade-baat ratio en zijn daarom de beste kandidaten voor onmiddellijke curatieve behandeling. Voor oudere mannen zijn deze ratio's aanzienlijk hoger sinds de negatieve effecten relatief minder dalen dan de voordelen, wat betekent dat het minder gunstig is voor oudere mannen om onmiddellijk te behandelen.

De antwoorden op de onderzoeksvragen en de implicaties ervan zijn besproken in **Hoofdstuk 10**. Ook de voornaamste conclusies en aanbevelingen voor toekomstig onderzoek en praktijk zijn hier gepresenteerd.

Ten eerste is besproken wat de impact van screening en behandeling is op de waargenomen incidentie en mortaliteit van prostaatkanker in de Amerikaanse bevolking en PLCO trial. De resultaten tonen aan dat de voor- en nadelen van PSA screening in de Amerikaanse populatie en in de PLCO trial waarschijnlijk verschillend zullen zijn van de effecten waargenomen in de ERSPC trial. Het is daarom belangrijk om in de toekomst de impact van de verschillen tussen de twee trials verder te analyseren, om zo duidelijke conclusies te kunnen treken over de doeltreffendheid van PSA screening en om te bepalen of de trials consistent bewijs geven met betrekking tot de doeltreffendheid van PSA screening.

Ten twee bespreken we de klinische resultaten, met name de voor- en nadelen van PSA screening en vroege behandeling. De ERSPC trial toonde aan dat, door te screenen met de PSA test, de prostaatkankersterfte kan worden verlaagd. Onze resultaten laten zien dat ook de bijbehorende negatieve impact aanzienlijk kan zijn. De omvang van de negatieve impact hangt echter af van hoe erg het voor het individu is om te leven met gediagnosticeerde en behandelde prostaatkanker en met de potentiële bijwerkingen. "Shared decision making" heeft daarom een belangrijke rol in het besluit om wel of niet te screenen of te behandelen voor prostaatkanker. De klinische resultaten gepresenteerd in het tweede deel van het proefschrift zijn belangrijke informatie voor dit besluitvormingsproces. Door het niet screenen en behandelen van mannen die niet bereid zijn de bijwerkingen van de behandeling te accepteren zou de negatieve impact van screenen kunnen worden gereduceerd.

Het is cruciaal om in toekomstig onderzoek te bepalen voor welke exacte leeftijden het meest gunstig is om te screenen en in het bijzonder wanneer te stoppen screenen. Ook een screening protocol met een beperkt aantal PSA testen zou een optie kunnen zijn met minder negatieve gevolgen. Een "active surveillance" protocol aanbieden in plaats van onmiddellijke curatieve behandeling aan patiënten met laag risico kanker is ook een veelbelovende strategie welke de negatieve gevolgen van screening kan beperken. Hoewel het nog jaren zal duren voordat de resultaten van de actieve surveillance trial en cohorten beschikbaar zijn, zullen deze belangrijke informatie opleveren met betrekking tot de veiligheid en werkzaamheid van deze strategie. Zodra deze resultaten beschikbaar zijn is het belangrijk de optimale active surveillance protocol en follow-up criteria vast te stellen.

Model appendix

In this appendix, first, a general overview of the model is given. Second, the quantification of the input parameters of the ERSPC model are presented. This model, which is our base model, simulates the ERSPC Rotterdam and ERSPC Göteborg trial population. In the different chapters in this thesis we described studies where different populations in different settings were modeled. To obtain the models for the different studies we adapted the ERSPC model. Therefore, some input parameters differ somewhat between chapters. In the final part of the appendix the differences between the models of the individual chapters are explained.

A.1 GENERAL MODEL OVERVIEW

MISCAN model is a micro-simulation model. For each individual in a population independent life-histories are generated including a possible cancer history, the effects of treatment and the effects of early detection by screening. The MISCAN-prostate model consists of four parts:

- Demography part
- Natural history part
- Treatment part
- Screening part

A.1.1 Demography part

The demography part simulates for each individual a life history without prostate cancer. For each individual, a date of birth and date of death from other causes than prostate cancer are simulated. A population can be built from different birth cohorts, where each cohort has its own birth tables and life tables. The distribution of births and deaths over calendar time can therefore be adjusted to represent the population under study.

A.1.2 Natural history part

The natural history part simulates prostate cancer histories for each individual separately. A certain proportion of the population might develop prostate cancer. The tumor starts in the preclinical phase. In the preclinical phase the tumor is asymptomatic, but can be detected by screening. From each preclinical stage, a tumor may grow to the next clinical T-stage (T1, impalpable; T2, palpable, confined to the prostate; T3+, palpable, with extensions beyond the prostatic capsule); it may dedifferentiate to a higher Gleason score (well differentiated, Gleason score 2-6; moderately differentiated, Gleason score 7; poorly differentiated, Gleason score 8-10); or it may be clinically diagnosed. For these transitions, the time spent in the current stage is generated from a Weibull distribution, where the parameters depend on the current stage and the next stage is determined by transition probabilities. In addition, there is a risk that a tumor in the local-regional stage will develop into distant disease, which is modeled by using a stage and Gleason score-specific hazard function. There is an additional

risk of being clinically diagnosed, which is modeled by using a stage-specific hazard function. This makes the model more flexible in the sense that it is possible to assume that the risk of clinical diagnosis is different for populations where the clinical practice might be different.

A.1.3 Treatment part

The life history after clinical diagnosis is determined by the treatment received. Different treatments can be assigned and the life history for individuals receiving a specific treatment is determined by treatment-specific survival of prostate cancer death. Treatment dissemination and survivals can depend on age, year, T-stage, metastatic stage and Gleason score at diagnosis.

A.1.4 Screening part

Screening is super-imposed on the life histories in the absence of screening. Screening tests applied to a person in a preclinical disease state may result in detection and alter his life history. A screening test is defined by its sensitivity. Screening is defined by the tests used, attendance rate and screening ages. Screening ages may be selected at regular intervals, or stochastically, allowing the modeling of both regular screening as in trials or screening programs and opportunistic screening. Screen detection of a tumor may alter the cause of events in an individual. We assume that the consequence of early detection by screening is that a percentage of the screen-detected men is cured from cancer and that for the other part detection does not alter the life history.

A.2 QUANTIFICATION OF INPUT PARAMETERS OF THE ERSPC MODEL

A.2.1 Demography part

In this model we simulated the ERSPC-Rotterdam and ERSPC-Göteborg trial population. The age at death from other causes is simulated per individual using Dutch life tables (Statistics Netherlands, 1991–1995). The birth tables were chosen such that the age distribution at the different screening rounds of the simulated trial populations met the observed age distributions.

A.2.2 Natural history part

The parameters of the natural history part were estimated by calibrating the ERSPC model to the data observed in the ERSPC-Rotterdam and Göteborg: baseline incidence (National Cancer Registry data for 1991[1]) and stage distribution in the Netherlands (Rotterdam cancer registry data 1992-1993[2]); baseline incidence in Sweden (1988-1992[3]); incidence, Gleason and stage distributions in the control arms of ERSPC-Rotterdam and Göteborg; and detection rates, interval cancer rates, Gleason and stage distributions in the screen arms of ERSPC-

Rotterdam[4-5] and Göteborg.[6] Number of cases diagnosed, and Gleason score and stage distributions in the control arms versus those in the screen arms give information about the disease progression through the various preclinical phases. Parameters were estimated by numerically minimizing the deviance between the number of cases observed and the number of cases predicted by the models. Deviances were calculated by assuming Poisson likelihood for incidence data and by assuming multinomial likelihood for stage-distribution data. The parameters for incidence, clinical diagnosis and sensitivity were assumed to be country-specific.

A.2.3 Treatment part

Treatment dissemination is modeled as a multinomial logit model with covariates age, T-stage and Gleason score at diagnosis. The categories of the multinomial model are radical prostatectomy, radiation therapy and active surveillance. The parameter estimates are based on data of the ERSPC trial section Rotterdam from the year 2000.

Survival curves for men with no initial treatment were estimated on the basis of SEER (Surveillance, Epidemiology and End Results) data, specifically on cases diagnosed in the pre-PSA era (1983-1986). The survival curves were modeled using Poisson regression with grade, stage, age and treatment type as explanatory variables. To assign the survival curves in our model we assumed that Gleason score 7 or less than 7 corresponds to grade well/moderately differentiated and that Gleason score more than 7 corresponds to grade poor/undifferentiated.

According to the results of Bill-Axelson et al,[7] we assumed that men receiving radical prostatectomy have a relative risk of 0.62 of dying from prostate cancer compared to men receiving no initial treatment. For men receiving radiation therapy we also assumed a relative risk of 0.62. For distant prostate cancer we assumed that treatment has no effect, implying that irrespective of the treatment type the survival is generated from the corresponding survival curve for men with no initial treatment.

A.2.4 Screening part

Depending on the frequency and sensitivity of the screening test, preclinical cancers may be detected by screening. Observed ERSPC Rotterdam and Göteborg data have been used to simulate the age and year specific attendance rate. PSA test and subsequent biopsy were modeled as a single test, where the sensitivity parameter was assumed to be clinical T-stage dependent. The sensitivity parameters were estimated simultaneously with the natural history parameters (A.2.2).

The effect of early detection by screening on mortality was included by assuming that a proportion of the screen-detected local-regional tumors are cured. The Gleason-specific cure rates were estimated by calibrating the ERSPC-Rotterdam model to the observed 27% prostate cancer mortality reduction in the overall ERSPC at follow-up after nine years.

A.3 MODEL DIFFERENCES FROM CHAPTER TO CHAPTER

A.3.1 Demography part

Chapter 2, 3 and 6

In Chapter 2, 3 and 6 we simulated the US population. In Chapter 2 we simulated the US population from 1975 to 2005, in Chapter 3 from 1985 to 2000 and in Chapter 6 from 1975 to 2000. The age at death from other causes was simulated per individual using US life tables from the Berkeley Mortality Database (http://demog.berkeley.edu/~bmd). The birth tables were chosen by generating a population with similar age distribution as in the core 9 registries of the Surveillance, Epidemiology and End Results (SEER).

Chapter 4

Same as the ERSPC model.

Chapter 5

In this model the PLCO trial population was simulated. The age at death from other causes was simulated per individual using US life tables from the Berkeley Mortality Database (http://demog.berkeley.edu/~bmd). The birth tables were chosen such that the simulated age distributions in each arm at randomization matched the observed age distributions.[8]

Chapter 7and 9

In Chapter 7 and 9 cohort analyses were done. The age at death from other causes was simulated per individual using Dutch life tables (Statistics Netherlands, 2000-2007). In Chapter 7 the birth cohorts were chosen such that the age of first screening was distributed uniformly from age 50 to 70. In Chapter 9 they were chosen such that the age of first screening was distributed uniformly from age 50 to 75.

Chapter 8

In this model the European Standard Population was simulated. The age at death from other causes was simulated per individual using Dutch life tables (Statistics Netherlands, 2000-2007). The birth tables were chosen such that the simulated age distributions in each arm at randomization matched the observed age distributions in the European Standard Population.[9]

A.3.2 Natural history part

Chapter 2, 3, 5 and 6

All parameters of the natural history part are assumed to have the estimates as calculated with the ERSPC model, except for the parameters defining the additional risk of being clinically diagnosed. This is because the clinical practice in the US population is possibly different than in the ERSPC trial. US-specific estimates of the hazard of being clinically diagnosed were obtained by calibrating the US model to observed age-specific incidence and stage-distribution (local/regional vs distant) in the US population. For this, data from the SEER registry of men aged 50 to 84 years who were diagnosed in 1975 to 2000 were used.

Chapter 4, 7 and 9

Same as in the ERSPC model.

Chapter 8

All parameters of the natural history part are assumed to have the estimates as calculated with the ERSPC model, except for the parameters defining the additional risk of being clinically diagnosed. Estimates of the hazard of being clinically diagnosed were obtained by calibrating the model to observed age-specific incidence in the Dutch population. For this, data from the Comprehensive Cancer Centers (http://www.ikcnet.nl) of men aged 55 to 74 years diagnosed in 1992 to 2002 were used.

A.3.3 Treatment part

Chapter 2

Treatment dissemination was modeled as a multinomial logit model with covariates age, year and grade at diagnosis. The categories of the multinomial model are radical prostatectomy, radiation therapy and active surveillance. Conditional on patient's characteristics available at diagnosis and primary therapy hormone therapy is assigned by a logistic model. Trends in primary treatments are based on data from SEER and trends in receipt of hormone therapy, adjuvant or neoadjuvant androgen deprivation therapy (ADT), are based on data from CaPSURE.

The survival curves for men with no initial treatment based on SEER data were used (see A.2.3). Also the same treatment efficacy parameter estimate (0.62) was used for men receiving radical prostatectomy or radiation therapy in combination with hormone therapy. For men receiving only radiation therapy we assumed the hazard ratio relative to no initial treatment to decrease from 0.9 in 1990 to 0.7 or 0.8 in 1995 and remain constant thereafter. Also several sensitivity analyses with different treatment efficacy estimates were performed.

Chapter 3 and 6

In Chapter 3 and 6 treatment was not modeled.

Chapter 4

Same as in the ERSPC model.

Chapter 5

In Chapter 5, in the PLCO model, it was assumed that 84% of men diagnosed with cancer received curative treatment and that 16% went into active surveillance. These are the percentages observed in the PLCO trial.[8]

The same survival curves for men with no initial treatment and the same treatment efficacy parameter estimate were used as in the base model (see A.2.3).

Chapter 7 and 9

In Chapter 7 and 9 it was assumed that all men diagnosed with cancer received curative treatment. In the chapters the same treatment efficacy parameter estimate (0.62) were used as in the base model. In Chapter 7 also the same survival curves for men with no initial treatment, which were based on SEER data, were used. However, these survival curves only distinguish between two Gleason score groups: Gleason score 7 or less than 7, and Gleason score more than 7. For the research question of Chapter 9 it was important to have at least three Gleason score categories and clinical T-stage as explanatory variables. Therefore, to obtain Gleason score and clinical T-stage specific survivals, we used the Albertsen[10] Gleason score specific survival curves and added clinical T-stage as an explanatory factor. For this we used the Cox proportional hazard estimates from Aus[11] (T1: 1, T2: 1.51, T3: 2.77) and we changed the Albertsen model such that the weighted sum of hazards by clinical T-stage add up to the same overall level by age and Gleason score.

Chapter 8

Treatment dissemination is modeled as a multinomial logit model with covariates age, T-stage and Gleason score at diagnosis. Treatment assignment for local-regional cases was based on the primary treatments (radiation therapy, radical prostatectomy and active surveillance) assigned in both arms of the ERSPC.

Survival of unscreened and untreated men diagnosed with local-regional prostate cancer was modelled using Albertsen Gleason score specific survival curves.[10] For distant disease, survival curves were based on SEER data.

The effects of treatment were modelled by assuming a relative risk of dying from prostate cancer of 0.65 for radical prostatectomy[12] compared with men receiving no initial treatment. This effect was also assumed for radiation therapy. For distant prostate cancer we assumed that treatment has no effect.

A.3.4 Screening part

Chapter 2, 3 and 6

For the dissemination of PSA screening in the US population we used the results of Mariotto et al,[13] who retrospectively constructed PSA screening histories in the population by use of survey data from the 2000 National Health Interview Survey and claims data from the linked SEER-Medicare database (http://healthservices.cancer.gov/seermedicare/).

US-specific estimates of test sensitivities were obtained by calibrating the US model (See A.3.2, Chapter 2, 3, 5 and 6) to observed age-specific incidence and stage-distribution (local/regional vs distant) in the US population. For this, data from the SEER registry of men aged 50 to 84 years who were diagnosed in 1975 to 2000 were used.

Chapter 4

Same as in the ERSPC model.

Chapter 5

In Chapter 5, in the PLCO model, we modeled the control arm and the screen arm. For the frequency of opportunistic PSA screening before and after the trial period we used the model simulating the dissemination of PSA screening in the US population (see A.3.4, Chapter 2, 3 and 6). Men randomized to the screen arm switch to annual PSA screening with random 85% compliance at each scheduled screen[8] for the trial intervention period then resume population screening. Men randomized to the control arm switch to a 20% higher intensity of screening than the general population during the trial to match previous estimates of PLCO control arm contamination (including reproducing the average 2.7 routine PSA tests during the trial[14]) then resume population screening.

Parameter estimates for the test-sensitivities were as estimated by calibrating the US model to the observed age-specific incidence and stage-distribution (local/regional vs distant) in the US population (see A.3.2, Chapter 2, 3, 5 and 6).

Parameter estimates for the effect of screening on mortality were estimated by calibrating the ERSPC model to the data observed in the ERSPC trial (see A.2.4).

Chapter 7 and 9

In Chapter 7 we assumed age of first screening was distributed uniformly from age 50 to 70 and subsequent annual screening until age 75. In Chapter 9 we assumed age of first screening was distributed uniformly from age 50 to 75 and subsequent screening every four year until age 75.

Parameter estimates for the test-sensitivities and the effect of screening on mortality were the same as estimated by calibrating the ERSPC model to the data observed in the ERSPC trial (see A.2.4).

Chapter 8

The following screening strategies were simulated: annual screening in the age groups 55-69 years and 55-74 years, screening at 4-year intervals between 55-69, and single screens performed either at age 55, 60 or 65 years. An 80% participation rate was assumed.

Parameter estimates for the test-sensitivities and the effect of screening on mortality were the same as estimated by calibrating the base model to the data observed in the ERSPC trial (see A.2.4).

REFERENCES

1. Visser O, Coeberg JW, Schouten LJ. Incidence of cancer in the Netherlands,1991. *Utrecht (The Netherlands): The Netherlands Cancer Registry.* 1994:13, 45.
2. Spapen SJ, Damhuis RA, Kirkels WJ. Trends in the curative treatment of localized prostate cancer after the introduction of prostate-specific antigen: data from the Rotterdam Cancer Registry. *BJU Int.* Mar 2000;85(4):474-480.
3. Parkin DM, Whelan SL, Ferlay J, Raymond L, Young J. *Cancer Incidence in Five Continents, Vol. VII* Lyon: IARC;1997.
4. 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.* Jun 18 2003;95(12):868-878.
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.* Mar 3 2010;102(5):352-355.
6. Hugosson J, Aus G, Lilja H, Lodding P, Pihl CG. Results of a randomized, population-based study of biennial screening using serum prostate-specific antigen measurement to detect prostate carcinoma. *Cancer.* Apr 1 2004;100(7):1397-1405.
7. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *New England Journal of Medicine.* May 5 2011;364(18):1708-1717.
8. Andriole GL, Grubb RL, 3rd, Buys SS, et al. Mortality results from a randomized prostate-cancer screening trial. *New England Journal of Medicine.* Mar 26 2009;360(13):1310-1319.
9. European Standard Population. Office for National Statistics, Newport, UK. (Accessed November 13, 2008 at http://www.statistics.gov.uk/StatBase/xsdataset.asp?vlnk=1260&More=Y)
10. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA.* May 4 2005;293(17):2095-2101.
11. Aus G, Robinson D, Rosell J, Sandblom G, Varenhorst E. Survival in prostate carcinoma--outcomes from a prospective, population-based cohort of 8887 men with up to 15 years of follow-up: results from three countries in the population-based National Prostate Cancer Registry of Sweden. *Cancer.* Mar 1 2005;103(5):943-951.
12. 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.* Aug 20 2008;100(16):1144-1154.
13. 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.* May 1 2007;109(9):1877-1886.
14. Pinsky PF, Black A, Kramer BS, Miller A, Prorok PC, Berg C. Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial. *Clin Trials.* 2010;7(4):303-311.

Dankwoord

DANKWOORD

Een proefschrift maak je niet alleen en daarom wil ik graag iedereen bedanken die op de een of andere manier heeft bijgedragen aan de totstandkoming van dit proefschrift. Een aantal wil ik hier persoonlijk noemen.

Allereerst wil ik mijn promotor, prof. de Koning, bedanken. Harry, bedankt voor het vertrouwen, de begeleiding en de ondersteuning, die onmisbaar waren in het hele traject. Jouw inzicht en deskundige kennis hebben veel bijgedragen aan dit proefschrift.

Gerrit, bedankt voor je begeleiding, ik heb veel van jou geleerd. Je bent al een tijdje met pensioen, maar ik kon altijd op je rekenen wanneer ik je expertise nodig had. Ik heb veel bewondering voor jouw kennis, maar ook voor je enthousiasme voor het vak. Jij hebt de basis prostaatkanker model in MISCAN opgebouwd en dat is iets om trots op te zijn. Dit is het eerste proefschrift dat gebaseerd is op dit model en hopelijk volgen er nog velen!

Eveline, bedankt voor de zeer fijne samenwerking! De afgelopen jaren was jij mijn eerste aanspreekpunt en ik heb veel aan je gehad. Ik vond het prettig dat ik met jou van gedachten kon wisselen over het werk. De gesprekken met jou hebben vaak geleid tot goede ideeën, die belangrijk waren voor de voortgang van de artikelen en dit proefschrift.

Leden van de kleine commissie, prof. Bangma, prof. Bindels en prof. Bossuyt, jullie wil ik van harte bedanken voor jullie bereidheid dit proefschrift te beoordelen.

Part of the work in this thesis was conducted in the prostate cancer CISNET-project. I would like to thank all people from the CISNET who have contributed to this thesis, especially, Roman, Alex, Ruth, Angela and Rocky. It has been a privilege to be able to work with you. I have learned a lot from our discussions and I think that the quality of this thesis is a result of the great collaboration. I would also like to thank Ann, Connie, Chen and Jing for always being so friendly to me at the CISNET meetings.

I would like to thank prof. Schröder, prof. Hugosson, prof. Bangma and Monique for their contribution to several articles presented in this thesis.

Ik wil graag iedereen van de afdeling maatschappelijke gezondheidszorg van het Erasmus MC bedanken. Ik wil in het bijzonder mijn kamergenoten bedanken voor alles wat we samen hebben meegemaakt. Maggie, het was fijn om met jou lief en leed te delen. Bedankt voor alle gezelligheid en steun. Irina en Marie-Louise, ik heb een ontzettend fijne en gezellige tijd gehad met jullie als kamergenoten. Bedankt voor jullie betrokkenheid en gezelligheid. Julie

en Moniek, al waren jullie voor een wat kortere tijd mijn kamergenoten, ook jullie bedankt voor deze fijne tijd.

Naast mijn kamergenoten wil ik ook andere collega's bedanken: Nicolien, Rianne, Kevin, Paula, James, Iris, Carlijn, Maaike, Marjolein, Joost, Jacque, Janneke, Serdal en Tiago; bedankt voor alle gezellige werkgerelateerde en niet-werkgerelateerde praatjes. Tiago, take good care of the model! Lifang, thank you for being a good friend. Liz, thank you for the nice talks and also for checking the introduction and discussion of this thesis. Stefan, thank you for being there whenever I needed advice. Farsia, bedankt voor je betrokkenheid en voor het regelen van mijn werkplek, kamergenoten en andere zaken, vooral in de laatste periode. Arry, bedankt dat je de screensectie zo goed ondersteunt.

Ik wil mijn studievrienden (Marjon, Jiehui, Hector, Marian, Hanneke, Susanne, Neslihan en Romeo) bedanken voor hun voortdurende interesse in mijn onderzoek en ook voor het bezorgen van de nodige afleiding.

Elisa, Estefhany en Jovanka, we kennen elkaar al jaren en het is mooi om te zien dat ook als alles blijft veranderen, ik altijd kan rekenen op jullie vriendschap. Elisa, we hebben al zoveel fases in ons leven samen meegemaakt. Ik vind het daarom leuk dat we allebeide de afgelopen tijd bezig waren met onze proefschriften. Wie had 25 jaar geleden, toen we speelden op de speelplaats van de Pire Kleuterschool, dit kunnen voorspellen?

Mama en papa, bedankt voor jullie onmisbare ondersteuning en liefde. Jan Gerard en Adriaan, lieve broers, bedankt dat ik altijd op jullie kan rekenen.

Lieve oma, bedankt voor al jouw liefde en alle mooie herinneringen. Je vond de ontwikkeling van je kinderen en kleinkinderen altijd heel erg belangrijk en ik weet daarom dat je enorm trots zou zijn geweest op dit proefschrift.

Lieve Zef, bedankt dat je er altijd voor mij bent. Ik ben zo blij met jou!

About the author

CURRICULUM VITAE

Elisabeth Marie Wever was born on the 17th of April 1984 in Curaçao. In 2002 she completed her secondary education (VWO) at the Radulphus College in Curaçao. In that same year she started with the study Econometrics and Operational Research at the University of Tilburg in the Netherlands. In 2007 she obtained her Masters Degree in Quantitative Finance and Actuarial Sciences. For her master thesis she did research on stochastic reserving methods for large claims. This research was based on an internship at SNS REAAL in Utrecht. Between 2007 and 2012 she was appointed as a researcher at the department of Public Health at the Erasmus MC in Rotterdam. During this period, she did research on the effects of prostate cancer screening and treatment using the microsimulation model MISCAN-prostate, which resulted in this thesis.

LIST OF PUBLICATIONS

Draisma G, Etzioni R, Tsodikov A, Mariotto AB, Wever EM, Gulati R, Feuer EJ, de Koning HJ. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *Journal of National Cancer Institute.* 2009;101(6):374-83. Doi: 10.1093/jnci/djp001.

Heijnsdijk EAM, der Kinderen A, Wever EM, Draisma G, Roobol MJ, de Koning HJ. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British Journal of Cancer.* 2009;101(11):1833-1838. Doi: 10.1038/sj.bjc.6605422.

Wever EM, Draisma G, Heijnsdijk EAM, Roobol MJ, Boer R, Otto SJ, de Koning HJ. Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. *Journal of National Cancer Institute.* 2010;102(5):352-355. Doi: 10.1093/jnci/djp533.

Gulati R, Wever EM, Tsodikov A, Penson DF, Inoue LYT, Katcher J, Lee SY, Heijnsdijk EAM, Draisma G, de Koning HJ, Etzioni R. What if I don't treat my PSA-detected prostate cancer? Answers from three natural history models. *Cancer Epidemiology, Biomarkers & Prevention*. 2010;20(5):740-50. Doi: 10.1158/1055-9965.EPI-10-0718.

Wever EM, Draisma G, Heijnsdijk EAM, de Koning HJ. How does early detection by screening affect disease progression? Modeling estimated benefits in prostate cancer screening. *Medical Decision Making*. 2011;31(4):550-558. Doi: 10.1177/0272989X10396717.

Gulati R, Tsodikov A, Wever EM, Mariotto AB, Heijnsdijk EAM, Katcher J, de Koning HJ, Etzioni R. The impact of PLCO control arm contamination on perceived PSA screening efficacy. *Cancer Causes & Control.* 2012;23(6):827-835. Doi: 10.1007/s10552-012-9951-8.

Etzioni R, Gulati R, Tsodikov A, Wever EM, Penson DF, Katcher J, Heijnsdijk EAM, Draisma G, Feuer EJ, de Koning HJ, Mariotto AB. The Prostate Cancer Conundrum Revisited: Treatment Changes and Prostate Cancer Mortality Declines. *Cancer*. Epub ahead of print, May 2012. Doi: 10.1002/cncr.27594.

Wever EM, Hugosson J, Heijnsdijk EAM, Bangma CH, Draisma G, de Koning HJ. To be screened or not to be screened? Modeling the consequences of PSA screening for the individual. *British Journal of Cancer.* Epub ahead of print, July 2012. Doi: 10.1038/bjc.2012.317.

Heijnsdijk EAM, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, Kwiatkowski M, Villers A, Páez A, Moss SM, Zappa M, Tammela TLJ, Mäkinen T, Carlsson S, Korfage IJ, Essink-Bot ML, Otto SJ, Draisma G, Bangma CH, Roobol MJ, Schröder FH and de Koning HJ. Quality-of-life effects of prostate-specific antigen screening. *New England Journal of Medicine*. 2012;367(7):595-605. Doi: 10.1056/NEJMoa12041637.

Wever EM, Heijnsdijk EAM, Draisma G, Bangma CH, Roobol MJ, Schröder FH, de Koning HJ. Treatment of PSA-detected prostate cancer: Benefits and harms by prognostic factors. *Submitted*.

PhD Portfolio

Name:	Elisabeth Wever
Erasmus MC Department:	Public Health
PhD period:	2007-2012
Promotor:	prof.dr. H.J de Koning
Supervision:	dr. G. Draisma
	dr. E.A.M. Heijnsdijk

	Year	Workload (ECTS)
General courses		
The why and how of readable articles	2007	0.6
Introduction to R	2008	0.9
English biomedical writing and communication	2010	4.0
Specific courses		
Erasmus Summer/Winter Programme, Erasmus MC, Rotterdam		
- Principles of research in medicine and epidemiology	2007	0.7
- Introduction to decision making in medicine	2007	0.7
- Survival analysis for clinicians	2008	1.4
- Topics in health and diseases in the elderly	2008	0.7
- Clinical trials	2008	0.7
- Principles of epidemiologic data-analysis	2009	1.4
- Case-control Studies	2010	0.7
- Markers and Prognostic Research	2010	0.7
Department of Biostatistics, Erasmus MC, Rotterdam		
- Multistate-models and models for competing risks	2009	0.6
- Bayesian statistics	2009	0.9
Karolinska Institute, Stockholm		
- Essentials of descriptive cancer epidemiology	2010	1.2
Nederlandse vereniging voor oncologie, Ellecom		
- Introduction into clinical and fundamental oncology	2011	1.4
Presentations		
Cancer Intervention and Surveillance Modeling Network (CISNET) meetings, National Cancer Institute, United States	2007-2012	6.0
Research meeting at the department of Public Health,	2008/2011	1.2

Erasmus MC, Rotterdam		
Methodology club at the department of Public Health, Erasmus MC, Rotterdam	2008	0.6
International conferences	2009/2011	1.2

Conferences

International Microsimulation Association, Ottawa	2009	0.9
Society for Medical Decision Making, Chicago	2011	0.6
Active Surveillance for Low Risk Prostate Cancer, Rotterdam	2012	0.6
Werkgroep Epidemiologisch Onderzoek Nederland, Rotterdam	2012	0.3

Seminars/Symposia

Seminars at the department of Public Health, Erasmus MC, Rotterdam	2007-2012	3.6
Cancer screening symposium: trial and modeling to guide public health policies, Rotterdam	2009	0.3
Netspar symposium: projective future life-expectancy, an interdisciplinary perspective, Rotterdam	2011	0.3

Teaching activities

Supervising 3rd year curriculum medical students for a community project, which is part of education theme 3.C 'Arts en volksgezondheid' at the Erasmus MC, Rotterdam	2011	0.8